DOCUMENT RESUME

ED 283 534 IR 052 006

AUTHOR	Holibaugh, Ralph
TITLE	Optical Discs for Storage and Access in ARL Libraries. SPEC KIT 133.
INSTITUTION	Association of Research Libraries, Washington, D.C. Office of Management Studies.
PUB DATE	Apr 87
NOTE	128p.
AVAILABLE FROM	Association of Research Libraries, Office of Management Studies, 1527 New Hampshire Ave., NW, Washington, DC 20036 ($20.00).
PUB TYPE	Collected Works - General (020) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Academic Libraries; *Access to Information; Higher Education; *Information Storage; Library Administration; Library Automation; *Library Planning; *Optical Disks; Pilot Projects; *Program Development; *Program Evaluation; Research Libraries; User Needs (Information)
IDENTIFIERS	*Association of Research Libraries

ABSTRACT

This SPEC (Systems and Procedures Exchange Center) kit provides a sample of excerpts from technical and planning documents contributed by 40 Association of Research Libraries (ARL) member libraries that are planning to install or have already implemented optical disc technology. Selected to represent a variety of media and applications, the excerpts address general as well as technical issues. Comments received in telephone interviews with contributors, as well as with other ARL members whose endeavors are not represented in this kit, are incorporated in a concise summary which discusses the overlapping and interrelated concerns of technical considerations, management issues, and impact on user services. In addition, the following original documents are included: (1) a descriptive list of optical disc technology terms from the National Library of Canada; (2) planning documents from the libraries of the Massachusetts Institute of Technology and the Smithsonian Institution; (3) proposals from Kent State University, the Massachusetts Institute of Technology, the University of Michigan, Syracuse University, and Vanderbilt University; (4) descriptions of pilot/demonstration projects from the Library of Congress and the Massachusetts Institute of Technology; and (5) evaluative reports from Brown University, the University of Illinois, and Vanderbilt University. A 10-item selected reading list and SPEC kit evaluation and order forms are also provided. (KM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED283534

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY
Maxine Sitts

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

System and Procedures Exchange Center

ORDER FORM

FED. ID #52-0784198-N

THIS ORDER FORM MAY BE USED AS A PROFORMA INVOICE

202-232-8656

***** KITS ARE AVAILABLE BY SUBSCRIPTION, FROM YOUR LIBRARY VENDOR/AGENCY OR DIRECT *****

To order, fill out below and send with **PREPAYMENT*** to: ARL Office of Management Studies, 1527 New Hampshire Ave. NW, Washington, DC 20036. Claims must be made within 90 days. (*ARL library members may be billed. ARL per-Kit cost is $10.00)

PREPAYMENT IS REQUIRED

Purchase Order #:____________________

Date: ___/___/___

OMS Client No.:__________
(see mailing label)

Shipping Address: ____________________
____
____
________________ (ZIP) ________

TOTAL NUMBER of Kits ordered: ________
X________ **$20 per Kit ($25 outside U.S.)**

TOTAL AMOUNT ENCLOSED $________

For Internal Use
Date ___/___/___
Postage

___ 133.Optical Discs
___ 132.Library-Scholar Communic.
___ 131.Collection Develop.Organiz.
___ 130.Retrospective Conversion
___ 129.Organization Charts
___ 128.Systems Office Organiz.
___ 127.Interlibrary Loan
___ 126.Automated Library Systems
___ 125.Tech.Services Cost Studies
___ 124.Barcoding of Collections
___ 123.Micro.Software Policies
___ 122.End-User Search Services
___ 121.Bibliographic Instruction
___ 120.Exhibits
___ 119.Catalog Maintenance Online
___ 118.Unionization
___ 117.Gifts and Exchange Function
___ 116.Organizing for Preservation
___ 115.Photocopy Services
___ 114.Binding Operations
___ 113.Preservation Education
___ 111.Coop.Collection Development
___ 110.Local Cataloging Policies
___ 109.Staff Training Automation
___ 108.Strategic Planning
___ 107.University Archives
___ 106.Electronic Mail
___ 105.Nonbibl. Databases
___ 104.Microcomputers
___ 103.Asst./Assc.Director Position
___ 102.Copyright Policies
___ 101.User Studies
___ 100.Collection Security
___ 099.Branch Libraries
___ 097.Building Renovation
___ 096.Online Catalogs
___ 095.Lib.Materials Cost Studies
___ 090.Integrated Lib.Info.Systm
___ 086.Professional Development
___ 085.Personnel Classif.Systems
___ 084.Public Serv.Goals-Objs.
___ 083.Approval Plans
___ 082.Document Delivery Systems
___ 081.Services to Disabled
___ 080.Specialty Positions
___ 079.Internships/Job Exchanges
___ 078.Recruitment-Selection
___ 076.Online Bibliog.Search Servs.
___ 075.Staff Development
___ 074.Fees for Services
___ 073.External User Servs.
___ 072.Executive Review
___ 070.Preservation Procedures
___ 068.AACR2 Implement.Studies
___ 067.Affirmative Action Progs.
___ 064.Indirect Cost Rates

SEE BACK

___ 063.Collective Bargaining
___ 057.Special Collections
___ 056.External Communication
___ 054.Internal Communication
___ 049.Use of Annual Reports
___ 048.External Fund Raising
___ 047.Automated Cataloging
___ 046.Planning Card Catalog
___ 031.Allocation of Resources
___ 029.The Systems Function
___ 020.Manager-Tech.Specialists
___ 019.Staff Associations
___ 008.Collective Bargaining
___ 007.Personnel Classif.Schemes
___ 003.Personnel Organization
___ 001.Organization Charts

USES OF SPEC KITS

The Systems and Procedures Exchange Center (SPEC) is a clearinghouse operated by the Association of Research Libraries, Office of Management Studies that provides a central source of timely information and materials on the management and operations of large academic and research libraries. It facilitates the exchange of knowledge and documents through SPEC Kits, which are distributed ten times each year to ARL members and other interested libraries. The Kits include topically-arranged groupings of unedited primary source documents – selected for their value to administrators and decision-makers – that illustrate a wide range of alternative approaches to specific issues.

Kit documents come from general membership surveys and from selected libraries contacted directly by SPEC, and most Kits are produced within six months of surveys. The documents' value comes from their variety of ideas, methods, and solutions. They are not viewed as finished products, but rather as points of departure for a library's planning efforts and as stimulants to innovative approaches to problem-solving. As such, Kits do not present answers or prescriptions for any one library; instead they illustrate how selected ARL members are planning for or dealing with particular issues. The worth of any one Kit to a particular library will depend upon the specific topic covered and the library's stage of development in that area.

Materials are selected according to the following criteria:

- Presents an approach of potential value to administrators and decision-makers
- Timely, and dealing directly with the topic under consideration
- Probability of application of ideas or thinking to other library situations
- Illustrative of actual practice, rather than theoretical
- Understandable, readable communication

All together, the materials should provide a range of alternative approaches that complement each other, provide variety, and stimulate comparison and contrast.

Libraries can take advantage of the Kit compilations in a number of ways. Administrators can evaluate the assumptions, methods, and results of other libraries' approaches; compare and contrast them; and use the learnings in their own situations. Library staff members can use the kits as professional development and current awareness tools. Committees and task forces can use them to begin a review of current practices. And the Kits can identify other persons or places to contact for further information. Back-up files in the SPEC office also are available for loan to member libraries. In addition, SPEC will conduct on-demand surveys or analyses geared specifically for a single library.

EVALUATION

Kit Title/Number ______________________________

1. Which uses did the library make of this Kit?

2. Please indicate how useful the Kit was for these purposes.

☐ Very Useful ☐ Quite Useful ☐ Somewhat Useful ☐ Not Useful

3. Do you have suggestions for this Kit or for future Kits?

(optional)

NAME ______________________________

LIBRARY ______________________________

PHONE ______________________________

Please return this form to the SPEC Center, OMS/ARL, 1527 New Hampshire Ave., N.W., Washington, DC 20036.

1/82

spec flyer
#133, April 1987
OMS
Systems and Procedures Exchange Center
ISSN 0160 3582

OPTICAL DISCS FOR STORAGE AND ACCESS IN ARL LIBRARIES

[EDITOR'S NOTE: In this flyer, optical disc is used as the generic name for videodiscs and digital media, including CD-ROM. The focus of the flyer is on alternatives for library planning and implementation of optical technology, rather than a technical discussion of the various media formats.]

The steady growth of videodiscs as image storage devices and, more recently, optical discs as mass storage media for textual, numeric, and graphic/image information, promises to address some basic storage and access problems facing research libraries. An increasing number of commercially available, library-oriented products, e.g., historic photographs on videodiscs or bibliographic databases on CD-ROM (Compact Disc-Read Only Memory), reflects optimism in the marketplace that this new technology will provide responses to these basic problems.

In pursuing projects which incorporate optical disc technology, several research libraries are moving toward integrating optical disc storage into new levels of service. What the proliferation of new technologies and alternative media carries with it, however, is a degree of risk in selecting appropriate formats. Before implementing the technologies, it is not only important to understand the capabilities of these formats, but to address attendant management issues as well.

The purpose of this SPEC Kit is to provide a sample of excerpts from technical and planning documents contributed by several members of the Association of Research Libraries (ARL) which are planning to install or have already implemented optical disc technology. The excerpts were selected to represent a variety of media and applications and to address general as well as technical issues. This flyer summarizes these documents and incorporates comments received in telephone interviews with contributors as well as other ARL members whose endeavors are not represented. Responses and material submitted by 40 libraries point to three overlapping and interrelated concerns: 1) technical considerations; 2) management issues; and 3) impact on user-services.

TECHNICAL CONSIDERATIONS. Some ARL members are quite optimistic about the potential of optical disc information storage media. General budgetary concerns, however, have prevented many research libraries from doing more than trying to keep up with the literature or experimenting with a few of the smaller, less expensive products. Some libraries are overcoming financial restraints by developing proposals for outside funding of optical disc projects. Pilot and demonstration projects enable libraries to gain experience with the technology as well as a sense of user responses.

Especially in its interactive mode, the videodisc in this decade is coming to be accepted as a viable instructional tool. Perhaps as a result of this, its use as a mass storage-retrieval device by research libraries, especially for color images and full-motion video, has grown noticeably in the past five years. The present state of optical media in libraries is much more volatile. The first kit document—a 1985 descriptive overview of optical disc formats—illustrates the complexity of the technology.

Despite widespread interest in optical discs in ARL libraries, extensive planning for or implementation of library applications continues to be limited to a few libraries within the federal government and those academic libraries receiving grant funding for projects having an optical component. The application of optical disc technology to data storage has been underway for some time, but only with the recent development of the CD-ROM has the cost become manageable for many research libraries. While enthusiasm was almost immediate for the CD-ROM's random access to large amounts of data, early marketing efforts were flawed by a lack of available disc drives. Software/hardware packages were sold in an attempt to address this problem and, in spite of some reservations about the cost advantage of bundled products, this practice continues.

The announcement of the CD-I (Compact Disc-Interactive) shortly after the CD-ROM emerged continues to affect library responses to CD-ROM. There has been a growing awareness of the appropriateness of image processing, full-motion video, graphics and multimedia in addressing information concerns. To some extent, CD-I follows the lead of the videodisc, and some librarians find it likely that after beginning as a consumer entertainment product, CD-I will later be developed as a medium for storing still images similar to the generic videodisc or for multimedia information products.

The brief history of CD-ROM has been plagued with problems of hardware/software compatibility. One of the first steps toward the resolution of these problems has yet to be completed, i.e., the establishment of standards. Only recently, through the efforts of the ad hoc industry committee known as the High Sierra Group, have standards for volume and file formats been proposed for adoption by the National Information Standards Organization. Proposed standards for CD-I call for two operating modes which will allow interchangeability with CD-ROM disc drives and media.

MANAGEMENT ISSUES. Research library administrators are increasingly responsible for choosing and integrating a variety of technologies into a unified approach to information delivery. The lack of effective interfaces between optical disc media and magnetic-based systems has been one of the biggest obstacles to more significant use of optical disc media thus far. Multi-function automated library systems are waiting for cost effective ways to integrate optical disc-based files, and with them a solution to the hardware and software problems associated with use by multiple users. Regardless of the impressive video or data storage capacities of optical disc media, many research libraries are not enthusiastic about the proliferation of discrete products or "systems," nor the possibility of single function workstations.

For research libraries with branches or multiple campuses, CD-ROM brings with it the issue of how optical media should be distributed. It is always difficult to justify the purchase of multiple copies of expensive hardware and media, but especially when the media are in easily transferred, portable, electronic form. On the other hand, planning for a central information facility using optical storage media may not always be cost effective. Such decisions benefit by the compilation of reliable data on the users of material in specific subject areas.

With deregulation of the telecommunications industry and the growth of satellite and fiber optics communications as well as optical disc media, some library administrators anticipate more decisions involving a "communication/mass storage trade off." Similar decisions may include balancing institutional mainframe storage and retrieval with optical disc enhanced microcomputer-based systems.

IMPACT ON SERVICE TO USERS. Within the context of a service organization, it is natural that many management issues are, in fact, also the central concerns of users. There is no single technology clearly superior in handling all facets of information delivery. It is evident that research libraries and their users are both tied ever more closely to technology, and that technology has become synonymous with change. Users (as well as library staff) who have become somewhat comfortable with using certain technologies in their search for information are now faced with learning how to operate even more advanced and varied systems. In addition to concerns with users' information seeking habits and behaviors, libraries also must deal with issues such as retrieval time, multi-function workstations, multiple users, document delivery and electronic remote access.

Research library users will continue to compare the operation of optical disc media with the established performance of the magnetic disk. Similarly, researchers will continue to use comprehensive, well-indexed online databases which are commercially available, and portable, easy to use information sources in hard copy as benchmarks for optical disc products.

User issues are often tied to software issues. Just as librarians and users are sensitive to hardware-specific information products, they are also aware of the shortcomings of file-specific search software. Users need retrieval routines in standardized, general purpose software that accommodate both magnetic and optical disc media. More precise studies of how users find and manipulate information are needed, in order for librarians to assist in the development of appropriate software. Important decisions about user services–determining the balance between professional, mediated service and direct access by the generalist–should be informed by such studies.

TRENDS AND NEEDS. With the rapid evolution of optical disc technology today, the task of determining which formats and products to use, and for which applications, is extremely complex. When the technology itself is in flux, many libraries decide to proceed cautiously with investments of time and financial resources. Yet there is the opportunity for some libraries to provide pioneering leadership for their colleagues. Regardless, this is a technology that deserves close monitoring.

Libraries were first drawn to CD-ROM because of the potential of its large data capacity, but experience thus far suggests that even larger capacities are often needed for mounting individual or related databases. CD-ROM jukebox devices with multiple drives may be in retrospect a diversion on the path to larger discs in libraries. A significant new technology is the optical digital disc, which—unlike CD-ROM—is a true computer peripheral storage medium. The new optical digital disc is a multi-user technology with six times the storage capacity of CD-ROM and faster response time. WORM (write-once read-many) and DRAW (direct-read-after-write) are two types of optical digital discs under development.

Meanwhile, many research libraries are waiting to evaluate features of write-once and erasable (rewritable) CDs. In addition to the obvious advantages of these media, they are also important in archival applications since the stability of present optical discs is estimated to be no greater than 10 years. Libraries also will be evaluating a variety of other anticipated digital media, including DataROM, CD-PROM (Compact Disc-Programmable Read Only Memory), and by 1991, CD-VI (Compact Disc-Video Interactive).

While optical disc media offer research libraries opportunities for enhanced or less expensive information storage, there is a need to temper market enthusiasm with critical comparisons using the current levels of service available with magnetic media. Mingling technologies and tapping larger amounts of data will require innovative software. The future of the videodisc as well as CD-ROM and CD-I, for example, will be affected by the development of software such as VideoDOS, which is compatible with CD-ROM and enables the creation of hybrid (analog and digital) videodisc applications. Additionally, smarter software, e.g., the intelligent access to optical disc databases by means of logic-programming languages, or recent work incorporating fuzzy Boolean searching, needs to be developed.

As the scope of administrative concerns expands from physical collections to include the information process, which is often not site specific or library centered, multi-function information workstations and telecommunications networking with local and remote databases should be worked into a unified information system approach. The mass storage opportunities of optical disc media need to be incorporated into a systems solution to the problems of information delivery as a whole.

The *Kit on Optical Disks (#133*, April 1987) contains 1 descriptive list of optical disc technology terms, 3 planning documents, 5 proposals, 2 descriptions of demonstration projects, 3 reports, and a selected reading list. (111 pages)

* * *

This flyer and kit was prepared by Ralph Holibaugh, Associate University Librarian, Fondren Library, Rice University, as part of the OMS Collaborative Research/Writing Program.

SPEC Flyer copyright 1987© by The Association of Research Libraries, Office of Management Studies. This SPEC Flyer may not be reproduced in any form without permission from the publisher.

SPEC Kits (ISSN 0160 3582) are available by subscription through subscriptions services or direct from the publisher. Individual issues cost $20.00, plus $5.00 postage outside the U.S., through distributors or direct. (ARL member price $10.00.) Prepayment is required. If ordering direct, send check payable to "ARL Office of Management Studies" to: SPEC, Office of Management Studies, 1527 New Hampshire Ave., N.W., Washington, D.C. 20036. Order by Kit Number and Title. Direct Order Forms listing back issues and Subscription Forms are available from the same address.

Table of Contents

Systems and Procedures Exchange Center

Optical Discs for Storage and Access
in ARL Libraries
Kit #133, April 1987

TECHNOLOGY AND APPLICATIONS

National Library of Canada — From "Canadian Network Paper," 1985. 1

PLANNING

Massachusetts Institute of Technology — "CD-ROM System as Back-Up to the Online Catalogue," September 1986. 8

Smithsonian Institution Libraries — "Report on The Use of Optical-Digital Disk as an Archival Storage Medium," selected pages, April 1986. 9

"SIL Optical Disk Working Group, Functional Requirements," selected pages, April 1986. .. 23

PROPOSALS

Kent State University — "Regional Campus Librarians Development of a Regional Campus Libraries' Database, A Proposal", selected pages, May 1986. 35

Massachusetts Institute of Technology — "A Proposal to Project Athena for a Videodisc Resource System," selected pages, January 1985. 46

University of Michigan — "BIBLIO: A Library-Based Academic's Workbench," selected pages, May, 1986. 50

Syracuse University — "The Kellog Project at Syracuse University," selected pages, 1986. 59

Vanderbilt University — "Enhancing Automated Information Services Within the Jean And Alexander Heard Library at Vanderbilt," selected pages, November 1985. 68

PILOT/DEMONSTRATION PROJECTS

Library of Congress
"Optical Disk Pilot Program: A Report on the Nonprint Project," December 1985. 73
"Library of Congress Optical Disk Pilot Program Comparative Characteristics of Analog Videodiscs as Used at the Library of Congress," February 1985. 77
"The Library of Congress Optical Disk Pilot Program," 1985. ... 79

Massachusetts Institute of Technology
One-Page Description of Barker Engineering Library Demonstration Site, February 1986. 81
"Set-Up Procedures for the CD-ROM System," February 1986. 82
"CD-ROM Searching in Barker Library," February 1986. 83
"Instructions For the Use of the CD-ROM Searching System," February 1986. 84
CD-ROM Bibliographic Searching System Evaluation Questionnaire, February 1986. 87

EVALUATIONS

Brown University
"The Brown University Library's Infotrac Experience," September 1986. 89

University of Illinois
"Creation and Demonstration of Optical Disk Version of the University of Illinois/River Bend Library System Online Catalog LSCA Project: III-86-0485-30 Final Report," selected pages February 1987. 90

Vanderbilt University
"Enhanced Information Access Project: Project Report #2," February 1987. 109

SELECTED READING LIST 111

Complete documents are available from the SPEC office.

Technology and Applications

10

Reprinted with permission of the National Library of Canada
Publication deadlines prevented inclusion of a 1987 revision of this document.

PART I:
THE TECHNOLOGY AND ITS APPLICATIONS

Roddy Duchesne
Senior Network Officer
Office for Network Development
National Library of Canada

1. OPTICAL DISK TECHNOLOGY

1.1 The Terms "Optical Disk" and "Videodisc"

The term "optical disk" in this report refers to a disk on which information is recorded in the form of markings (pits, bumps, spots or marks) by an optical system. The optical recording system normally employs a laser. A common width of marking is 0.5 microns.[3]

The term "videodisc" in this report refers to a disk which:

a) is encoded in an analogue format for replay on a TV or TV-compatible monitor;

b) employs a similar recording density in bits per unit area to optical disk; and

c) is designed to be read using either an optical or electrical capacitance (non-optical) detection system.

The basic difference between analogue and digital encoding is noted in the next section, which reviews the main types of disks. The term "digital videodisc" is sometimes seen in the literature, but is not used in this report. Digital videodisc refers to analogue videodisc used to store digital data. Digital/analogue conversion is used with such disks. The term "analogue optical data disk" is used to refer to these disks in this report. Examples of these disks are LaserData Inc. LaserDiscs and disks used by the Reference Technology Inc. DataDrive 2000 system.

Following the above definitions, developments such as optical tape and optical cards are not covered in this report.

1.2 Types of Optical Disks

Optical disks and videodiscs are categorized in numerous ways. Some of the principal categorizations are:

- digital/analogue
- optical/capacitance
- read-write; read-only; write-once; erasable
- direct read after write/pressed
- reflective/transmissive
- diameter size.

These categorizations are reviewed briefly below:

a) Digital/analogue

Digital disks employ digital encoding. Conceptually, the encodable spiral on the disk is a set of regularly spaced binary bits. Each of the regular spaces either contains a marking or it does not, corresponding to binary 1 or binary 0. Markings and spaces between markings are in fact longer towards the circumference of the disk and shorter towards the centre of the disk. Binary digits or bits may be used to represent information in a variety of ways. Characters may be encoded in eight bit bytes, for example, or in an image (facsimile) form. In image form, the image is divided into pixels (picture elements). Depending upon factors such as the need to represent shades of grey, partial colour spectra or full colour spectra, there may be one bit or as many as 128 bits per pixel.[4]

Analogue disks employ pulse code modulation (PCM) that represents a continuously variable (analogue) FM signal encoded by variances in the laser marking length and the space between markings.[5]

Analogue encoding is employed by consumer-market disks designed for playback of audiovisual materials, such as feature films, on TV or TV-compatible monitors. Analogue encoding is also employed by the LaserData Inc. DataDisc system, and the Reference Technology Inc. DataDrive 2000 system. Both systems can be used to store and retrieve digital data in addition to analogue data and both use digital/analogue conversion to record digital data (for example, digital computer data) on analogue disks.

b) Optical/capacitance

An optical disk is recorded using a light source, normally a laser. A similar light source, normally a lower-powered laser, is used for reading by current optical disk systems.

Capacitance systems employing magnetic encoding are not currently on the North American new-equipment market. Production of the only capacitance players sold in North America was discontinued in the spring of 1984; these were the Capacitance Electronic Disk (CED) players marketed principally by RCA. The other major capacitance system, called Video High Density (VHD), is marketed in Japan and England but not in North America. There is a possibility that VHD may be introduced into the American interactive industrial market within a year or so. It is not expected to be introduced into North American consumer markets.[6] Both CED and VHD systems employ electrical capacitance writing and reading systems.

A hybrid optical/magnetic or magneto-optic system has been developed by the Japanese company, Daicel

Chemical. This system employs a laser which heats up spots on the disk, producing a reversible magnetic field which is then read by a capacitance system.

c) Read-write; read-only; write-once; erasable

Read-write systems can both read disks and write on them. An example of such a system is the Alcatel Thomson Gigadisc[7] system outlined in Table 3.

Read-only systems can only read disks. Except for the Alcatel Thomson Gigadisc system, the other systems in Table 1 and Table 2 are of this type.

Write-once is a term used in contrast to erasable. The Alcatel Thomson Gigadisc system in Table 3 is a write-once system. Erasable systems are at the research and development and prototype stage. There are still performance and reliability problems with erasable media.[8]

d) Direct read after write/pressed

These terms refer to the means of producing the disk. Direct read after write (DRAW) is a process in which information is written on a disk a track or track sector at a time. "Track" is used here in the sense of one complete 360 degree turn of the read/write spiral on the disk. Immediately after writing, the information is read back and checked for errors. In the event of errors, writing of the same information is attempted in the next track or sector. The process is repeated until a track or sector is successfully written with no detected errors.

Pressing is a process used to replicate disks from a suitable master. The process is similar to that used to replicate phonograph records; it is quick, cheap and used to mass-produce disks. Quantity-produced pressed disks have a much lower unit cost than direct read after write disks, which are generally lower in unit price for single disks or disks in very small quantities. The March 1985 Pioneer mastering charge for single-sided linear CAV (Constant Angular Velocity) videodiscs was $US2100, and $US4200 for two-sided disks. Replication was $US10 for each single-sided disk and $US15 for each double-sided disk for the first 1000 disks; these prices dropped, respectively, to $US7 and $US12 for quantities over 1000. Interactive CAV and all CLV (Constant Linear Velocity) videodisc prices were from 15 percent to 25 percent higher. In mid-1984, Philips Megadoc blank DRAW disks were quoted at $US600 per single-sided and $US900 per double-sided disk.

Error correction techniques are used to achieve acceptable corrected BER (bit error rate) in both direct read after write and pressed disks. Both processes are used to produce disks with corrected BER of better than one undetected bit in error in 10^{12} bits. Achieving higher degrees of freedom from error entails the use of more disk space for parity and error correction bits and redundancy associated with error correction. This disk space is not available for user data and is often referred to as "storage overhead." In LaserData Inc. LaserDiscs, about 30 percent of total disk capacity is taken up with error correction and track/sector addressing.[9] In Philips Megadoc disks, some 5 percent of total capacity is used for header information and another 20 percent is reserved for error correction.[10]

e) Reflective/transmissive

Reflective disks have a mirror-like finish. If the disk is encoded on both sides, it must be turned over for reading the second side. Transmissive disks do not need to be turned over to read the second side since the reading laser can be refocussed to read the second side through the disks, which are made of translucent material. In spite of this advantage, the vast majority of optical disks are reflective disks.

f) Diameter size

The most common disk diameter sizes are 4.72 inches (12 cm) and 12 inches (30 cm). These are nominal sizes and not necessarily the exact measurements of disks made by particular manufacturers. The smaller size is used by the widely marketed Compact Digital Audio Disc (CDAD) system. The larger size is used by consumer and industrial market videodisc systems and is used by a number of optical data disk systems.

There are a number of other disk sizes. The five nominal disk sizes for which recorded and unrecorded standards proposals have been submitted are 14, 12, 8, 5.25 and 4.72 (12 cm) inches.[11] Eight-inch optical reflective videodiscs are currently marketed in North America, mostly containing rock music and video program material. The Japanese Video High Density system uses ten-inch diameter capacitance disks.

1.3 Major Types of Currently Marketed Optical Disk Systems

Table 2 and Table 3 below note the salient characteristics of the four principal types of optical disk systems on the North American commercial market: Compact Digital Audio Disc (CDAD), videodisc, digital optical data disk, and analogue optical data disk. These system types are reviewed briefly below.

1.3.1 Compact Digital Audio Disc (CDAD)

These 4.72-inch diameter disks are strictly standardized, and are also commonly referred to as Compact Disc or sometimes as Audiodisc. CDADs are used as a medium for very high quality audio and are mass-produced for the consumer market.

Currently marketed disks contain only audio, no video or computer-readable data. CDADs are produced by a pressing process and are optical reflective and read-only.

CDAD player makes and models differ mainly in features such as the number of selections which can be entered into memory and played back in a pre-selected order. There are 25 brands of players on the North

American market, including Fujitsu, Kenwood, Panasonic, Philips, Matsushita, Mitsubishi, Sanyo, and Sony.

The National Library of Canada's Public Services Branch has an Alpine 7200 player. Features of this player include:

- 16-program random memory function allowing the storage in memory of up to 16 tracks/disk manufacturer index numbers and the playback of these in any designated order;
- uncued fast forward/fast backward at 75 times normal speed; and
- scanning at ten times normal speed with cue sound playing at normal pitch for the first ten seconds of each track/indexed item.

Table 1 below compares CDAD and conventional LP system characteristics. Key CDAD advantages are:

- superior sound quality;
- given initial quality of the master tapes from which the sound is copied; much less noise; "A hi-fi enthusiast's dream come true";[12]
- virtual freedom from wear and tear, and from dust and cleaning problems associated with conventional records;
- much less susceptibility to vibration: for example, there are car models on the market;
- remote control and random access features; and
- longer maximum playing time – approximately three times longer per side; in practice, CDADs are currently marketed with a single playing side and this is not fully used; this generally results in a single-sided CDAD having the same playing time as a double-sided LP.

1.3.2 Videodisc

The format for videodisc is not as standardized as that for CDADs. Several very different formats are in use in North America. The largest number of disks sold to date are in the 12-inch diameter CED (Capacitance Electronic Disk) format used by the now discontinued non-optical CED system marketed principally by RCA. Since over half a million CED players were sold, and 12 million CED disks were sold in the year preceding RCA's decision to cease production, there may be a continuing market for these disks even though players are no longer being manufactured.[13]

RCA's decision to discontinue production of CED players appears to spell the demise of the CED system and leaves optical reflective videodisc as the major videodisc format. Major manufacturers of optical reflective videodisc players include Pioneer, Sony, Hitachi and Philips. The National Library of Canada has a Pioneer PR7820 machine which it used in its Videodisc Demonstration Project (reported in Part 2 of this report). The Pioneer LDB-6000 player noted in Table 2 is a later model, with 2K of built-in memory compared with the 1K of built-in memory in the PR7820. The Library of Congress's operational Analog Video Disk System uses Sony disks and players.

Optical reflective videodisc players of one make can generally play disks made for another make of player, provided that play is straight-through and not interactive. Interactive play involves the reading of programs from the disk and their execution by the microprocessor in/with the player. It is possible to master an interactive disk which will include programming to run on different makes of players. However, disks are normally mastered for one make of player. Interactive features can be used on the make of player for which the disk was mastered, but may not work on other makes of players.

TABLE 1

Comparison of Compact Digital Audio Disc and Conventional LP System Characteristics[14]

	Compact Digital Audio Disc (CDAD) System	Conventional LP Player
Specifications		
Frequency response	20 Hz - 20 kHz ± 0.5 dB	30 Hz - 20 kHz ± 3 dB
Dynamic range	More than 95 dB	75 dB (at 1 kHz)
S/N	95 dB	60 dB
Harmonic distortion	Less than 0.01%	1 - 2%
Separation	More than 90 dB	25 - 35 dB
Wow & flutter	Quartz precision	0.03 %
Dimensions		
Disk diameter	4.72 in./12 cm	12 in./30 cm
Playing time: minutes per side	60 (max. 74)	20 - 25
Operation/Reliability		
Durability - Disk	Semi-permanent	High frequency response degraded after being played several tens of times
- Stylus	Over 5000 hours	500 - 600 hours
Operation	• Quick and easy access due to micro-computer control • A variety of programmed play possible • Increased resistance to external vibration	• Needs stylus pressure adjustment • Easily affected by external vibration
Maintenance	Dust, scratches, and fingerprints are made almost insignificant	Dust and scratches cause noise

Twelve-inch diameter pressed, read-only, analogue, optical reflective videodiscs are on both the North American consumer and institutional/industrial markets. Mass-produced consumer disks are used mostly for replay of feature film, general education and how-to material (for example, how to play tennis). Institutional/industrial market disks are commonly used for instructional purposes (for example, technician training); for promotional/publicity purposes (for example, US Army recruiting); entertainment (for example, arcade games); and for storage and retrieval of audiovisual materials (for example, art images).

Eight-inch diameter optical reflective videodiscs on the North American consumer market are used mostly for rock music and video. These disks can be played on the same players as 12-inch diameter optical reflective videodiscs.

TABLE 2

Illustrative Compact Digital Audio Disc and Videodisc Systems

	Compact Digital Audio Disc (CDAD)	Videodisc
System used as an example	Alpine 7200 with remote hand control	Pioneer LD-V6000 with remote control unit
Disk diameter	4.72 in./12 cm	12 in./30 cm
Disk characteristics	digital; optical reflective; read-only; replicated by pressing process	analogue; optical reflective; read-only; replicated by pressing process
Two-sided disks?	feasible, but only single-sided disks marketed in mid-1984	yes
Playing time: minutes per side	74 maximum; most commercial disks play 45	30 (CAV) 60 (CLV)
Disk rpm	200-500 (CLV)	1800 (CAV) 600-1800 (CLV)
Disk life in years	>10	>10
Mid-1984 Pre-recorded disk price range: $Canadian	15-25	25-40
Mid - 1984 Player price range: $Canadian	1080	2050
Access time in seconds with correct disk side mounted	2 seconds to selected track	maximum 3 seconds to specified frame
Dimensions W x H x D mm inches	435 x 81 x 332.5 17.1 x 3.2 x 13.1	420 x 150 x 415 16.6 x 5.9 x 16.4
Weight kg lb	6.9 15.2	14 30.8
Light source	laser diode	laser diode
Operate in normal office environment	yes	yes
Additional equipment required to operate		
a) minimum	a) amplifier; pair of loud-speakers	a) colour monitor
b) desirable additions for enhanced performance		b) amplifier; pair of loud-speakers
Main competing systems	LP phonograph records; audio tape systems	videocassette systems

Videodiscs can be produced in CAV (Constant Angular Velocity) or CLV (Constant Linear Velocity) format. In the CAV format, one video frame is recorded per track, or – more strictly – per 360 degree turn of the videodisc. This format provides random access to each frame and sharp freeze frame, but has a lower storage capacity than the CLV format.

With the CLV format, one video frame per track is recorded on the inner portions of the disk, expanding to three frames per track on the outer portions of the disk. Only limited random access is normally possible with CLV disks, and no freeze-frame, but CLV disks can hold twice as much data as CAV disks. Twelve-inch diameter CAV videodiscs store up to 54 000 NTSC (United States National Television System Committee standard) video frames per side with accompanying stereo sound. Twelve-inch diameter CLV videodiscs store up to 108 000 NTSC video frames per side with accompanying stereo sound. Playing time for 54 000 NTSC frames is 30 minutes, played straight through at 30 frames per second.

NTSC frames provide a resolution of approximately 50 x 50 lpi (lines per inch), not generally suitable for recording whole pages of text. As noted in Part 2, an average of four NTSC frames per page was found to be required for the legible recording and playback of 8 1/2 x 11 inch pages of text.

The most nearly competitive system to videodisc is videocassette. Compared with videocassette systems, optical reflective videodisc systems have a number of advantages:

- much greater play-life;
- fast random access;
- for CAV format disks: random access to each frame on the disk and a sharp still frame which can be maintained over long periods without wear;
- superior picture quality to even high quality, high speed videocassette – reproduction of this quality on the monitor depends on the quality of the monitor;
- superior sound quality similar to hi-fi videocassette – reproduction of this quality depends upon amplifier and speaker arrangements;
- non-erasability – for applications where this is an advantage; and
- technical characteristics (for example, random access and still frame) much better suited for interactive computer-assisted applications such as interactive learning programs.

Compared to videocassette systems, optical reflective videodisc systems also have a number of disadvantages:

- disks cannot be erased and re-recorded;
- higher playback cost per hour: VHS SLP tapes with a three-hour or more playback time have a playback cost per hour of $US3 or less[15] compared to a CLV consumer-market optical reflective videodisc with a playback cost per hour of $US8 or less;
- much fewer available prerecorded titles – about 380 titles in March 1982, compared to 5000 BETA/VHS titles;[16] new titles have been added in both formats, but a very considerable disparity remains; and
- generally higher player cost, at least at the lower end of the price spectrum, since bottom-of-the-line VCRs are considerably cheaper than bottom-of-the-line videodisc players.

1.3.3 Digital Optical Data Disk

Optical Data Disk is frequently abbreviated to ODD. The acronyms OD^3 or OD3 are sometimes used to refer to Digital Optical Data Disk.

This type of disk is produced by a direct read after write process and may be read using a read-write unit or a read-only unit. Alcatel Thomson Gigadisc and Philips market 12-inch diameter optical reflective systems. Storage Technology Corporation employs 14-inch diameter optical reflective disks.

Key advantages of digital ODD systems are:

- Storage capacity

They are capable of handling very large quantities of data using mechanical disk changers ("jukeboxes"). The Philips Megadoc automatic disk exchanger handles up to 64 double-sided disks (i.e., 128 gigabytes), changing disks in about 15 seconds. At the time of writing, up to three exchangers could be handled by a single Megadoc system. Byte storage costs are much lower than magnetic disk; competitive with some microforms; higher than microfilm,[17] lower than on-line magnetic tape;[18] and becoming lower than magnetic tape on the shelf.[19] International Data Corporation quoted in 1984 the following storage costs per on-line megabyte: conventional tape drives $US150; IBM 3380 disk drives $US40; four gigabyte optical disk drive $US5.[20] LINK Resources Corporation has forecast that optical storage will have a serious impact on micrographic storage by the early 1990s.[21]

- Suitability of data rates for high speed data transfer

A critical factor in computer applications is the transfer of large files of machine-readable data, and high speed facsimile transfer. The Thomson Gigadisc data transfer rate is 3.83 Mbits/sec.

TABLE 3

Illustrative Optical Data Disk (ODD) Systems

	Digital ODD	Analogue ODD
System used as an example	Alcatel Thomson Gigadisc Model GD 1001/20 with controller GC 1001/10	LaserData Trio 110 controller; Pioneer LD-V1000 LaserDisc Player
Disk diameter	12 in./30 cm	12 in./30 cm
Disk characteristics	digital; optical reflective; read-write; recorded using direct read after write process	analogue; optical reflective; read-only; replicated using pressing process
Two-sided disks?	yes	yes
Playing time: minutes per side	35	30 (CAV)
Disk rpm	1122	1800 (CAV)
Disk life in years		
- before recording	5	10 year life or better normally quoted
- after recording	10	
- physical copy	10	
Approximate capacity per side:		
- user data bits	8×10^9(1GB)	6.4×10^9 (800MB)
Corrected bit error rate	10^{-12}	10^{-12}
Player access time with correct disk side mounted	within 40 track band - 5 msec; beyond band - 200 msec	1.3 seconds average (LD-V1000 player)
Operate in normal office environment	yes	yes
Additional equipment required to operate	computer and peripherals depending on application	microcomputer; microcomputer interface; monitor; peripherals according to application
Main competing systems		
- computer peripherals	magnetic tape; magnetic disk	magnetic tape; magnetic disk
- document/image storage	analogue ODD; microforms; magnetic tape paper	digital ODD; microforms; magnetic tape paper

- Ability to handle colour, limited only by input and output devices
- Suitability for applications requiring only a small number of copies of each disk

Digital ODD may also be pressed in quantity for electronic publishing in a similar manner to the way that Compact Digital Audio Discs are pressed in quantity.

- Reliability

Optical disk equipment is much less prone to mechanical failure (for example, head-crashes) than magnetic disk or tape. Optical disks are themselves much less sensitive to dust, wear and tear, and handling.

- Long lifespan

Alcatel Thomson Gigadisc quotes a ten-year lifetime after recording. A disk write-life of 2.5 years and a read-life of 7.5 years have been quoted for the Philips Megadoc system.[22] Magnetic computer tapes require refreshing every 20 to 24 months.

- Designed to operate in a normal office environment, unlike high-speed mainframe computer magnetic tape and magnetic disk peripherals

Digital Optical Data Disk systems have a number of limitations, often more associated with the marketing concept employed than inherent in the technology; for example, the Philips Megadoc document digitizer available in mid-1984 was designed to handle documents with a page size no larger than A4/8 1/2 x 11 inches; it also reduced colour images to black-and-white pixels. Some limitations common to current digital ODD systems are:

- Write-once

With present commercially available systems, once a read-write system has written on a portion of disk, that portion cannot be reused for new data. For some applications this is an advantage. Commercially available, operationally proven erasable optical disk systems are not expected before 1986.

- Equipment cost

Present digital ODD systems can be expensive when all necessary peripherals are included [for example, document digitizer; flying spot scanner; mini/microcomputer and software; high resolution monitor(s); audio input and output equipment; printer(s); automatic disk exchanger(s)]. In mid-1984 the Philips Megadoc System was priced at $US0.25 million to $US1.4 million, depending on the configuration.[23] Digital ODD systems depend on the application but are generally more costly in straight equipment cost than videodisc systems. The introduction of CDROM digital ODD systems will go a long way toward closing the gap in player costs between digital ODD and videodisc at the low end of the digital ODD equipment price spectrum.

- Slow speed of input digitizers

In mid-1984, the Megadoc image scanner was manually fed and digitized up to 1080 pages an hour; in practice, manual feeding is unlikely to reach and maintain this input rate. Micrographics equipment routinely converts 4000 or more pages an hour.

- Slow access time compared to magnetic disk

In a comparison made by GEAC Computer Corporation, the following figures were obtained:[24]

	Total read time msec	Total write time msec
Thomson-CSF GD 1001 digital optical disk	180	230
Magnetic disk	40	56

Speed of access is particularly important in accessing indexes to data bases in large multi-user multi-task mainframe computer systems. By contrast, access times in the one-to-three second range may be quite acceptable in small single-task microcomputer systems.

1.3.4 Analogue Optical Data Disk

This type of disk is essentially a conventional videodisc written in a manner which allows the storage of digital data via digital-to-analogue conversion and the inclusion of error-correction data. The marketing of systems using this type of disk is a recent development, pioneered in 1983 by LaserData Inc., Woburn, Massachusetts. LaserData Inc. LaserDiscs are optical reflective, read-only disks produced using a pressing process. Deliveries commenced in 1984. The newness of this system means that it is still proving itself and details of aspects such as pricing are more likely to change with it than with systems that have been on the market longer. Table 3 provides basic details. A system using the same basic technology is marketed by Reference Technology Inc., Boulder, Colorado. Reference Technology Inc.'s DataDrive 2000 system differs from the LaserData system in a number of ways. At $US8900 for quantity 100, the system was higher priced and offered faster access times than the LaserData system in mid-1984. A corrected byte (sic) error rate of 10^{-13} bytes was also claimed compared to the LaserData corrected bit error rate of 10^{-12}.[25]

According to the company's president Martin Hensel, the LaserData system is oriented towards "publishers of reference works, texts, directories, and periodicals who would market databases on videodiscs and equipment for using them."[26] The Reference Technology Inc. DataDrive 2000 system is expected "to be used to distribute databases that require large storage capacity and/or fast access times. These would include legal references, indexed catalogs, medical expert systems, scientific/technical information, library support, economics/econometrics, financial services, and government publications, which include all of the others. Many of these now need to be partitioned before they are distributed because no media has sufficient capacity to hold them. Such databases would benefit from the DataDrive 2000's 1GB capacity."[27]

The principal economic and performance differences between the LaserData/Reference Technology Inc. analogue ODD systems and the Alcatel Thomson Gigadisc digital ODD system are as follows.

The LaserData/Reference Technology Inc. system disk prices favour applications where disks are produced in quantity, for example, electronic publishing applications. They disfavour applications requiring only a very small number of copies of each disk, such as single institution preservation/archival applications.

The LaserData system employs a comparatively inexpensive and slow access time disk player. This tends to favour the use of the LaserData system for applications with numbers of user workstations where access times in seconds are quite acceptable. An example of this type of system would be a local reference workstation controlled by a microcomputer and providing access to audiovisual and/or textual data bases published in quantity on LaserDisc. The same parameters disfavour the LaserData system as a peripheral for a large computer system where speed of access is a critical factor. The Reference Technology Inc. DataDrive 2000 system is targeted to fill a market position between the LaserData system and the Alcatel Thomson Gigadisc digital ODD system. "The DataDrive 2000 offers performance comparable to a high end Winchester and ought to support up to about 16 users fairly easily." [28]

1.4 Future Developments

Some of the most dramatic developments are expected in the Compact Disc area. Some of these developments are:

* Large expansion in the number of CDAD titles

By August 1984, slightly more than 1000 titles had been released in CDAD format. This will at least double by the end of 1985, and continue to grow at a very rapid rate. At the same time CDAD prices will fall.

* Introduction of multiple CD products

JVC already had a 130 disk jukebox for CDs in mid-1984 priced at 600 000 yen ($US2500) and providing an average access time of 13 seconds. This time includes time to reject, search and begin play.[29] CDROM (Compact Disk Read-Only Memory) devices are expected within the first half of 1985. The capacity of a single CD is around 560 MB compared to an IBM PC diskette's 360KB. CDROM device prices have been predicted to drop from the area of $US1500 as production volumes increase and the market develops. CDs including graphic information (for example, song lyrics, album liner notes) are already on sale in Japan.[30] Later developments include CD-DRAW (CD-direct read after write).

Further developments involving disk formats other than CD, as well as CD are predicted:

* Combination CD/videodisc players

Pioneer Video of Montvale, New Jersey, has announced that it will market a combination Compact Disc/videodisc player. The player will use a common disk tray and will accommodate a variety of disk sizes, including 4.72-inch, 8-inch and 12-inch disks. It should be available in the United States in February 1985.[31]

* Computer peripherals

Optical data disks will become widely used as a computer peripheral in all types of computer installations. They will partly replace magnetic tape as a data archiving medium. ODD devices will also figure in Local Area Network (LAN) configurations. GEAC Computer Corporation Limited has already demonstrated a LAN (Net/One) linking Alcatel Thomson Gigadisc with a number of GEAC computers and terminals.

* Electronic publishing

Use of CD and 12-inch ODD for electronic publishing will increase, bringing the facility of making local access to very high use data bases, reference works in machine-readable form, and audiovisual materials to the local institution. There will also be increased use of optical data disk (in sizes including compact disks and 12-inch disks) for applications involving large quantities of data with very little updating required, such as data archives.

* Erasable disk systems

The commercial introduction, installation and use of erasable disk systems, probably at relatively high prices, by the end of 1986. This will be for commercial purposes rather than for home installation and use, since prices will be too high for the home market. Daicel Chemical has announced that it will mass-produce erasable disks around the middle of 1986.[32]

2. APPLICATIONS AND PROJECTS

A very large number of applications and projects are relevant to libraries. This chapter provides notes on operational applications and a few selected projects described as far as possible using cited published accounts. Any snapshot is quickly outdated: the purpose of the descriptions is not to provide up-to-date details, but to illustrate the types of ways that libraries and agencies concerned with information and document delivery are currently using the technology. For up-to-date details, the reader is referred to the project institutions, their publications, and the literature. For descriptions of a wider set of projects, the reader is referred to the "Videodisc Projects Directory" entries contained in successive issues of the journal *Videodisc and Optical Disk*, and proceedings of relevant conferences. *Optical Memory News* normally publishes a conference calendar in each issue.

2.1 Operational Applications

Operational applications in libraries in 1984 included:

- the use of videodiscs and CDADs as regular library materials made accessible to library patrons;
- the use of Digital Optical Data Disk in a library catalogue card storage, retrieval and printing system;

Planning

CD-ROM SYSTEM AS BACK-UP TO THE ONLINE CATALOGUE

A back-up to the catalogue will be necessary in the event of planned or unplanned downtime. We intend to implement a system for this purpose which emulates most closely our online catalogue and is least disruptive to our patrons and staff. Therefore, we have selected a back-up system utilizing CD-ROM (Compact Disc - Read Only Memory), a relatively new but very suitable technology for libraries. The back-up system would be accessed from the same workstations as the online catalogue, and would offer similar access to our bibliographic records. The CD-ROM players could also be used for other purposes, such as access to reference sources or bibliographic instruction, when not needed for the online catalogue. The project would involve joint development with a vendor of retrieval software to emulate our Geac interface; this software would be transportable to other libraries with Geac online catalogues.

The traditional medium for a back-up to the online catalogue has been computer output microfiche (COM). CD-ROM has several advantages over COM as an online catalogue back-up. The data on CD-ROM is machine-readable and can be randomly accessed, and the retrieval procedure is similar to that of the online catalogue. Data can be easily printed or downloaded into local files. The disks are much easier to use, maintain, and to distribute to remote locations. Data on a disk can also be distributed electronically over a network. Past experience has shown that COM back-up listings are not widely consulted, even during downtime; we are convinced that CD-ROM will be more successful for this function.

MIT has the expertise to accomplish this project. Library staff have experience with using and designing search retrieval software and in dealing with vendors on the development of new systems. The Catalogue Department is exploring the use of a CD-ROM system for a segment of retrospective conversion. The Barker Engineering Library has been serving as a demonstration site for the Digital Equipment Corporation's database distribution and end-user searching on CD-ROM. This has given the staff experience with the medium and insight into user reactions and the problems and/or advantages of particular search retrieval software. The MIT campus, and especially Project Athena, is very interested in new applications for CD-ROM, as compact disk players will shortly be plentiful. Distribution of our data to campus networks via CD-ROM would be highly attractive. This represents an exciting opportunity to work on the development of a new product to the benefit of both MIT and the wider library community.

ACTION:

- Select a vendor to produce the disks and work with the vendor on retrieval software, screen displays and search commands to emulate our Geac software.
- Purchase CD-ROM players and install them in our public areas.
- Generate a MARC tape of our database to send to the vendor to have pressed.
- Establish ongoing procedures for disk production and access.
- Establish a method of distribution of the disks to the campus community.

ATTACHMENT 1

Report on

The Use of Optical-Digital Disk as an Archival Storage Medium

Introduction

Optical-digital disk systems, with raster scanning as the input device, potentially offer researchers a new technology which combines on-line search capability with enhancement of images, manipulation of data, and publication quality printing. As this technology continues to develop and becomes less costly, it may also prove useful as an alternative system to microfilm for permanent capturing and storing archives. It already appears to be significantly advantageous for reference that does not require exacting standards of imagery.

As optical-digital disk systems began to develop in a laboratory environment for use in archives and libraries within the United States and Canada, the Smithsonian Institution Libraries established a working group to look into the possibility of using an optical-digital disk system in SIL. The Smithsonian Institution Archives (SIA), the National Museum of American History Archives Center, and the Office of Information Resources Management were asked to join the group and assist in investigating the potential of optical-digital disk systems as an archival storage system and as a quick access research tool.

Very early in the working group's investigations into existing systems it became apparent there were serious questions about the use of this technology as a permanent archival system. However, it became evident that these questions had little or no bearing on the use of the technology in a reference system, designed to speed access and reduce the handling of original documents.

Based on what is currently available, SIA recommends that the Smithsonian develop an optical-digital disk capability for use in a reference system. This service would enable researchers to gain access to information, while extending the life of original and fragile documents by alleviating the problems inherent with their constant handling. It would also safeguard against the loss of information through wear, theft, or misplacement.

In this type of system, current available technology would enable documents to be scanned at 400 pixels per inch (PPI), with a capability of providing hardcopy at 400 PPI. As technology improves, the Institution will require a scanner which provides optimum resolution at a reasonable cost.

The following study points out, however, that at the present stage of optical-digital technology, the use of raster scanning (input) and the storage of information on optical-digital disk, for permanent archival use, is cost prohibitive. This does not mean the technology should be ignored. Cost for scanning and capturing data decreases each year, and it is possible, within one or two decades that it will cost as little to store archival quality data on optical-digital disk as it now does to store information for reference copy use. For now, SIA firmly advises that archival quality micrographics is the only acceptable permanent archival storage alternative to original documents. Current, inexpensive, off-the-shelf optical-digital disk technology does not meet the needs for archival quality copying, even though it is advantageous for a reference system.

Archival Standards

It is assumed in any archival-quality copying system that the document or item to be copied is of permanent value, the item and its information will be needed indefinitely, and that any copy intended to be a security copy must meet the most exacting standards as a suitable substitute. This applies to both casual reference and for more demanding requirements of precise legal and scientific examination for incontrovertable evidence. Furthermore, the copy produced must be, in both its physical property and quality of image, durable indefinitely. It must maintain its original condition without appreciable diminishing of quality for "more than one hundred years," the generally accepted standard for "indefinite storage."

Such quality is obtained only by applying the most exacting and rigorous standards throughout the entire copying process: materials, equipment, procedures, and all other variables that may affect the quality and durability of the product. This drives up the cost of archival copying far beyond what is acceptable for most commercial purposes and even for most general information storage and transmission purposes, to the extent that this high cost discourages institutional investment in archival copying. It follows, therefore, that the technology to be applied must not only meet high standards, but it must do so in ways that will reduce costs substantially so as to offset the high cost of quality control and to encourage institutional investment.

Microphotography

Currently, silver-halide film, manufactured, processed, and stored according to national industry and the Association of Image and Information Management (formerly the National Micrographics Association [NMA]) standards, is considered the only accepted archival medium. Roll microfilm is relatively inexpensive to produce, takes less space than paper records, and has been accepted in lieu of original archival documents for reference use and in courts of law for tort and litigation actions, even though originals are preferred.

In a micrographic image copy, quality is measured in terms of resolution and density. Resolution is the ability to record fine detail, and its measurement is usually expressed as the maximum number of line-pairs per millimeter the film system can satisfactorily resolve.

The notion that mere readability should determine resolution is not acceptable as an archival standard. Readability is subjective. It is an unreliable test for determining what can be, and what has been captured from the original document. Textual information of poor quality takes longer to read and will not serve all purposes. Since image resolution can never be greater than the original copy, quality is best met by striving to reproduce the resolution available in the original document, rather than by attempting to estimate an acceptable loss of resolution that results from an arbitrary resolution standard.

Using the NMA MS23-1979 Quality Index graph for determining resolution, SIA requires that its microfilm system capture and resolve a lower-case letter "e" one millimeter in height, so that its legibility is of excellent quality up to and including the third-generation microfilm copy. This quality requirement, at a reduction setting of 12:1, means that SIA must resolve no less than 120 optical lines per millimeter, using the National Bureau of Standards resolution test chart, 1010A, on its master, or first generation camera copy.

This is equivalent to 3000 line pairs per inch, 6000 pixels (picture elements, see under "Raster Scanning," below, for definition of pixels) per linear inch on the film, or 500 lines per inch on the original. In addition, Nyquist's sampling theorem, which relates input to output resolution, requires a minimum of a factor of two increase in input resolution, to 12,000 pixels per linear inch on the film, or 1,000 pixels per linear inch on the original, to meet this requirement.[1] Because the microfilming system currently in use easily exceeds 120 lines per millimeter, to provide comparable quality to current archival microfilm publication, an archival imaging system must also comfortably exceed a resolution of 120 lines per millimeter.

Density is the light-absorbing or light-reflecting characteristic of a photographic image. In a microfilm system, the greatest concern is with the background density, measured in the background areas of the microfilm document. Controlling this background density provides optimum print quality with good contrast between textual material and background.

According to NMA MS23-1979, gross background densities from 0.70 to 1.50 in clear-base, negative apppearing film are recommended. This range represents five different categories of documents, from faint pencil lines on worn and faded paper to computer output microfilm and typed and published documents. In considering these recommendations, SIA assumes that: first, on any given reel of film produced, it will include documents representing four of these categories; and second, in copies produced from the master film, density tends to diffuse and splay. As a result, SIA has established a required density range of 0.90 to 1.20 for most of its microfilming projects.

In the process of achieving a background density of 0.70 to 1.50, a microfilm system can also accurately reproduce densities from 0.00 to 1.50 and beyond (computer generated microfilm can easily produce a density of 3.00). Thus, while not explicitly striving to reproduce the full density range, microfilm systems do achieve a faithful reproduction of the wide range of densities required to reproduce holographic documents as well as the watermarks, stamps, and marginalia that help to identify many printed documents.

---

[1] 120 optical lines = 120 line pairs = 120 x 2 ruled lines = 240 ruled lines = 240 pixels per millimeter on the microphotograph = 240 x 25.4 (mm per inch) = 6096 PPI on the digital photo micrograph = 6096 x (2^2 pixels minimum Nyquist synthetic aperature ratio) = 12191 PPI.

Raster Scanning, Optical-Digital Disk Systems

Assuming optical-digital disks have the physical capability and software algorithm to store archival information indefinitely, (given the ability of error correcting codes [ECC] to insure exact digital duplicates of optical disks up to the date they expire), the next question would be the quality of the image stored. In digital systems, resolution is measured in pixels (picture elements) per linear inch (PPI). As in microphotography, once the scanning (filming) has been completed, the resolution of the scanned document is fixed, and cannot be increased.

In raster scanning, the term used to quantify the brightness or darkness reflected from an image is grey scale. The number of bits per pixel determines the number of grey scale values. Normally, 8 bits of grey scale are recorded, matching the density range of microfilm, from 0.00 to 2.50, in 0.01 increments. It is this range of grey scale, or density, that allows microfilm to reproduce handwritten marginalia, stamps, and watermarks.

To reduce the amount of information stored in research quality optical-digital disks systems, the 8 bit grey scale is thresholded to one bit. In photography this is called silhouetting. As in photographic silhouetting, thresholding eliminates the tonal range required to reproduce handwritten marginalia, stamps, and watermarks.

To further reduce the amount of information or bits stored in an optical-digital disk system an algorithm which allows for the compression of data redundancy is used. For archival quality copy, the compressed algorithm must be isomorphic, or allow for the exact reconstruction of an image when the data is decompressed for reviewing.

Ideally, in an archival optical-digital disk system, all documents should be digitized at 2500 PPI, 8 bits per pixel for black and white, or 32 bits per pixel for the three primary colors plus black. The digital image should then be compressed, eliminating redundant data and storing only the amount of data required to reproduce the true resolution of the original document. At this setting all half-tones and color images are fully resolved.

Not all original documents need to be digitized at 2500 PPI, 32 bits per pixel (see Appendix II). However, this setting provides the maximum resolution required for most documents, and it is unlikely that it would be cost effective to change the resolution for each individual document. Documents that can be resolved at a lower setting will not significantly increase the number of unnecessary bits of information in the system. This is because these documents will have greater data redundancy which will be reduced by the data compression algorithm.

Cost Scenario.

In order to effectively compare the costs for storing archival quality digital images, a cost comparison is given between archival quality microfilm, archival quality images in a digital system, and reference use quality copy images in a digital system. In this scenario, costs for preparation, quality control, and other manual operations for capturing images have been excluded.

Certain elements in the analysis have been simplified and standardized. It is assumed that all documents within SIA have to be captured and stored in the various systems, all original documents are black and white (8 bits per pixel), of one size and form; that the compression ratio will remain the same regardless of document type (for optical-digital disk), and that each reel of microfilm will contain the same number of images.

To arrive at the cost of providing archival quality microfilm for all of the SIA's holdings, the following equation is used:

a. total volume of records (measured in cubic feet) x
b. number of pages in each cubic foot divided by
c. the number of pages on each reel of microfilm x
d. the cost to purchase raw stock film (in dollars) x
e. the cost to develop a reel of film (in dollars) equals
f. the cost for archival quality storage on film.

a. 10722 x
b. 2000 x
c. 1/700 x
d. 12 x
e. 5 =
f. $1,838,057 to store 21.444 million documents.

Based on this equation, it would cost $1.8 million to store 21.4 million documents, or $0.0857 per document.

To arrive at the cost of storing archival quality images (bits) through raster scanning on optical-digital disk the following equation is used:

g. total volume of records (measured in cubic feet) x
h. number of pages per cubic foot x
i. size of page (in inches) width x length x
j. pixels per inch squared x
k. bits per pixel for grey scale divided by
l. the compression ratio divided by
m. the number of bits in each byte divided by
n. the number of bytes in a gigabyte x
o. the cost of storing one gigabyte (in dollars) equals
p. the cost for archival quality storage of digital images.

g. 10722 x
h. 2000 x
i. 8 x 11 x
j. 2500 x 2500 x
k. 8 x
l. 1/10 x
m. 1/8 x
n. 1/1 billion x
o. 100 =
p. $117,942,000 to store 1.1 quadrillon bytes of archival images.

For archival quality imaging at 2500 PPI and 8 bits per pixel for grey scale it would cost $117.9 million to store 1.1 quadrillon bytes of information, or $5.50 per document.

The following calculations use the above equation to calculate the cost of storing a reference quality copy of the Smithsonian Archives in digital form:

g. 10722 x
h. 2000 x
i. 8 x 11 x
j. 400 x 400 x
k. 1 x
l. 1/10 x
m. 1/8 x
n. 1/1 billion x
o. 100 =
p. $377,414 to store 3.7 trillion bytes of information.

For a research copy at 400 PPI and 1 bit per pixel for grey scale it would cost $377,414 to store 3.7 trillion bytes of information, or $.0175 per document.

Conclusion

Optical-digital disk systems offer easier access to information, and are cost effective for research quality systems. While a research quality system is an important adjunct to any archival operation for the purpose of reducing physical access to documents, optical-digital disk systems are cost prohibitive at this time as an archival quality copy storage system. Even when comparing current digital storage cost of archival quality copies to the alternative of keeping original documents in off-site remote storage, having a staff person dedicated to retrieving the material, along with a vehicle and other equipment needed to carry out the retrieval process (all of which would not cost more than $150,000 per year, for the first year), it would take SIA 786 years before it would realize any cost benefit. However, assuming that the cost of digital systems will decrease every two years at a rate similiar to the last forty years, 1945-1985, the cost of an archival quality system will be equal to the current cost of the research quality system within twenty years, making digital storage a cost effective method for archival preservation of documents.

Appendix I. Glossary

Archival Record Film.	A photographic film composed and treated (washed properly to remove thiosulfate ions) so that under archival storage conditions it is suitable for the preservation of records having permanent value.
Bits Per Pixel	The number of bits it takes to store the grey scale of each pixel. For the conversion value, see grey scale, below.
Contrast Enhancement.	This is based on the worst case scenario which assumes the viewer can only distinguish twenty shades of grey. On a grey scale with 256 shades of grey, such a viewer cannot distinguish shades that are closer than 1/20 of 256 or 12.8 gradation. Stretching the gamma is a method of contrast enhancement that accommodates this inability. This method systematically changes the grey scale of all pixels in an image so that pixels of interest have a resulting gradation separation of more than 12.8 and are therefore distinguishable by the worst case viewer. However, because the difference between a symbol, a line, and the background level in a poor to low contrast document is available only if all 256 grey scale levels are recorded, stretching the gamma would only be effective on images for which at least eight bits of grey scale have been recorded.
Data Compression.	An algorithm allowing data redundancy to be removed. In a high quality document, with limited density variation, data compression is very effective. This is particularly true for publications and typewritten documents, where the type is uniform against a uniform background. For

holographs, written in pen and pencil, where the pressure applied to the instrument used for marking varies, for documents with stamps or watermarks, or where there is "noise" in the document, such as smudge marks and erasures, compression is less effective and the compression factor will be reduced.

Density. Microfilm image density is a measure of the modulation (attenuation) of light (or other spectral radiation) when the microfilm is illuminated on one side and viewed on the other. The measured value of density is the logarithm, to the base 10, of the ratio of radiant flux incident on the sample to the radiant flux transmitted.

According to NMA MS23-1979, gross background densities from 0.70 to 1.50 in clear-base, negative-appearing film are recommended. There are five categories listed in this standard, representing five different types of documentation, each with a density range at which the documents can be microfilmed sucessfully.

Grey Scale. In raster scanning, the term used to quantify the brightness or darkness of light reflected from an image. The number of bits per pixel (BPP) determines the number of grey scale values.

Bits per Pixel	Number of Grey Scale Values
1	2
2	4
3	8
4	16
5	32
6	64
7	128
8	256
9	512
10	1024
11	2048
12	4096

	One bit per pixel, representing two levels of grey scale, would equal black and white. The human eye can distinguish between twenty to thirty levels of grey scale. For archival quality copying, 256 levels of grey scale are required. This allows the viewer, by manipulating the gamma, to enhance a hard to read portion of a document.
Isomorphic.	For archival quality copy, the ability of the compression algorithm to exactly reconstruct the original digital image when the data is decompressed for reviewing.
Optical-Digital Disk.	A storage device for information in digital form as required for computer processing of data and facsimile images.
Pixels Per Inch.	Pixel (picture element) is the fundamental unit that makes up the image in a raster digitized document. The number of pixels within a given area determines the resolution of the digital image. Pixel resolution can be specified as pixels per linear measure (pixels per inch, PPI), or pixels per image, as in 525 line television.
Quality Index.	A subjective relationship between legibility and the resolving power pattern in a test chart. It can be used to ensure a particular level of quality or legibility in a micrographic system. The desired quality level is established by relating the height of the lower case "e," the number of generations in the system and a resolving pattern meeting the requirements of ANSI/ISO 3334, such as the NBS microcopy Resolution Test Chart 1010A.
Raster Scanner.	A method for converting the image on a document into computer code. In raster scanning an imaginery microscopic grid

	is placed over a document forming the raster. Each square of the grid contains an individual picture element. The intensity of light falling on each picture element, an analog value, is measured and conveyed as an analog signal, which in turn, is converted into a digital bit stream representing the lightness and darkness of areas found on the original document.
Resolution.	The ability to record fine detail. In digital systems, resolution is measured in PPI (picture elements per linear inch). Archival quality copying is defined as quality that can replace the original. The requirement for archival quality resolution is best met by striving to reproduce the resolution available in the original document rather than by attempting to estimate an acceptable loss of resolution that results from an arbitrary resolution standard. In microfilm, resolution is measured in optical lines (line pairs) per millimeter as specified in ANSI/ISO 3334.
Resolving Power.	A numeric expression of the ability to distinguish or separate two entities spaced closely together. In micrographics, it is the product of the number on the National Bureau of Standards 1010A test pattern resolved in the image multiplied by the reduction of the image and is expressed in line pairs per millimeter. Microfilm cameras are capable of recording 150 to 200 line pairs per millimeter on the film. So that the film does not constrain the camera (will accept everything the camera can record) microfilm can record 600 to 800 line pairs per millimeter. Nature provides a limit of 2000 line pairs per millimeter, which is imposed by the wave length of light to which microfilm is sensitive.

Appendix II. Document Resolution and Storage Cost Per Image in an Archival Digital System

These are the expected maximum resolutions required. The compressed volume of data for each type of image, when scanned at 2500 PPI, 32 bits per pixel, is expected to be less than the data volumes listed in the "cost scenario" equations above. The "cost for storage" figure, below, includes a factor of 10:1 data compression for each image. As in the previous cost scenario, it excludes preparation of information, quality control, and other attendant manual operations.

Type of Document	Size	PPI	BPP	Cost for Storage
Architectural Drawings	38x24			
original, black and white		1000	8	$ 9.12
original, color		1000	32	$ 36.48
reproduced, black and white		1000	8	$ 9.12
reproduced, color		1000	32	$ 36.48
Maps	46x35			
copper plates, color		1000	32	$ 64.10
holograph, color		1000	32	$ 64.10
holograph, black and white		1000	8	$ 16.02
printed from a black and white photograph	4x5	5000	8	$ 5.00
printed from a color photograph	4x5	2500	32	$ 5.00
Newspapers	14x24	2500	1	$ 2.62
Legal-Size Documents	8x14			
handwritten		1000	8	$ 1.12
typed		1500	1	$ .31
type-set		5200	1	$ 3.78
copy from an electrostatic copier		725	1	$ .07
Photographs (negatives)	4x5			
black and white		5000	8	$ 5.00
color		2500	32	$ 5.00

Type of Documents	Size	PPI	BPP	Cost for Storage
Posters	20x32			
original, color		1000	32	$ 25.60
reproduction	(see under photographs, above)			
Quarto Documents	8x11			
handwritten		1000	8	$ .88
typed		1500	1	$ .25
type-set		5200	1	$ 2.97
copy from an electrostatic copier		725	1	$ .06

These resolution requirement estimates are intended to be conservative. It might be possible to relax some or all of them after testing. However, digital archives is a new technology, and these estimates may prove not to be as stringent as needed.

III. A Pilot Project

The Optical Disk Working Group proposes a pilot project to convert portions of the S.I. Libraries Trade Literature Collection to optical digital disk. One of the Libraries' richest and most valuable resources, this collection documents the products and objects that contribute to the cultural, commercial and industrial history of the U.S. Consisting of over 300,000 pieces, the catalogs are of varied sizes and shapes, bound and looseleaf, whole books and single sheets. They contain black and white and color photographs, line drawings, colored plates and other illustrations. Paper quality varies from heavy coated stock to the yellowest newsprint. Typefaces range from miniature to poster size.

The Working Group feels that the trade literature collection offers an ideal body of material to be captured in a pilot project. It constitutes a homogeneous subject area, while at the same time offering a wide variety of image data to be captured. In addition, SIL is entering records for the trade catalogs into SIBIS, so that bibliographic access is available. Storing the catalogs in a system that had direct links to the SIL Online Catalog would provide a real service to all researchers using SIBIS, would preserve a unique resource, and would allow for a thorough testing of the technology.

IV. Scenario for Conversion of Documents to Optical Disk

This scenario is neither rigorous nor final but is designed to give the vendor a context in which to understand the specifications which follow. The Working Group expects the exact flow of work described here to be modified as the project proceeds.

A. Preparation of Documents

It is anticipated that material to be converted to disk would be brought to a central site for preparation and associated record-keeping. Some items may require disbinding; this would be done under the supervision of the Book Conservation Laboratory. Input staff would follow procedures similar to those used in preparing materials for microfilming, i.e. collating the material, preparing bibliographic and other forms, assigning control numbers, etc. To date we have found no way of reducing the staff costs required for this step. The input unit staff would be non-professional (i.e. GS 3-5) support staff; this would include the staff to operate the scanner.

B. Capturing and Indexing Documents

Upon receiving document(s) to be scanned, an input staff person finds the bibliographic record of the document. If one is lacking, the staff person creates a brief skeletal record. Through a simple command the bibliographic record number is linked to all the images captured for that document. The document is then loaded into the automatic feed mechanism of a scanner. If the document is not suitable for automatic feed then the operator enters it page by page. For each image scanned the system assigns a unique image number linked to the bibliographic record for the document. At some point in the input process, a staff person may indicate to the system the start of certain sections of the document such as table of contents, index, bibliographies, body of the document,etc. This is to help a researcher in retrieving those portions of the document.

The image is scanned and stored temporarily on a magnetic storage medium so that it can be displayed and reviewed. The operator can adjust for gross errors by manipulating a lightness/darkness setting similar to the lightness/darkness setting on any office copier. Additional image enhancement techniques can be applied at this point if necessary.

After scanning, an operator reviews the image and related indexing, coding etc. to verify the input quality, and write the document to an optical disk. If necessary for adequate review the operator can print out the image at a printer capable of duplicating the resolution at which the image was captured.

If some pages require rescanning or the data in them needs to be modified, they can be identified and rescanned or otherwise changed, whereupon they will replace the rejected images in the display.

V. Scenario for Use of System by Researcher

An information seeker would enter a branch library or archives unit and proceed to the SIBIS terminal to search the online catalog. He would find a citation to a trade catalog, which would carry a note such as "available on optical disk." The user would then proceed to the optical disk work station. Or, if the user knew the materials sought were likely to be contained on optical disk, he would go directly to the high-resolution terminal work station. In this case the researcher would enter the online catalog through the high resolution terminal, and locate the item in question in the same way as if it were a normal SIBIS terminal. Alternatively the researcher could scan the document from the normal low-resolution terminal. In any case the user should be able to request the display of the document without having to leave the online catalog software and log on to another system. The user might wish to go back and forth between the image display and the bibliographic citation.

Let us say that the researcher finds several citations to trade catalogs on refrigerators. Each bears the note "available on optical disk" with a number. The user selects the item from the menu, requests a display, and is automatically shown a menu displaying the sections into which the document has been logically divided, e.g. table of contents, text, index, bibliography, etc. If he wishes to go to any of these sections directly he can enter his selection from the menu. Otherwise, if he hits a "send" key, he is positioned at the first page/image of the document. By using "forward" and "back" commands, the user pages through the document. The user controls the speed at which he is browsing to reach the desired portion of the document. If he finds the necessary information, he stops; otherwise he requests a display of the second document from the menu. There he finds an engraving of a refrigerator and a description of its mechanical parts. Using the same print commands that apply to the online catalog, he requests a print of the page displayed, which comes from the workstation printer. If he wishes the entire document, he asks for a "batch print", which is performed by a centralized, high-speed printer and delivered to the branch. If he wishes, he may include a print of the bibliographic citation along with his printout. If he wants to review the catalog itself for any reason, he requests that the branch retrieve it from remote storage.

If the user then decides to consult refrigerator advertisements that would normally be found in an archival unit, he stays at the terminal and enters the archives online file. Browsing through the index he locates a "folder" of advertisements for kitchen appliances. He selects the folder from the menu, and requests a display of the first "page", i.e., the first page of the items in the folder. Using the browse function he continues through the folder, occasionally requesting prints. He takes the prints away to use in determining the items to be included tentatively in an exhibit on refrigeration technology.

VI. Specifications

(M is mandatory, D is desired)

A. Access

Here are included all means of retrieving the stored data. Data is used here in a broad sense including text stored as ASCII characters, text stored as images, and illustrations.

System will support retrieval of images:

M 1. Directly, without going through the bibliographic system. This might be used for modification, deletion, error checking etc., using a unique accession number referring to the "unit" of data (image, page, screen however defined). The accession number concept here implies a unique identifier, assigned by the system at time of input, which can be used to track the "unit of data" by staff. This is similar to "grsn" concept in SIBIS except instead of referring to a bibliographic record it refers to an image.

M 2. Through document number

M 3. Through automatic retrieval of images linked to a bibliographic record.

B. Link with online catalog

M 1. If there is a bibliographic record for the document in the online catalog, then at the time of scanning an operator will be able to establish a system link between the images for the document and the bibliographic record by wanding the barcode of the document or by entry of the bibliographic record number.

M 2. If there is no bibliographic record for the document in the online catalog at time of scanning, the system should allow for the creation of skeletal bibliographic record that may be linked to the images created.

M 3. User will be able to use the online catalog to retrieve data stored on optical disk. For example, user finds a citation, note, or indicator in an online catalog display of the record. Based on this, user tells system he wishes to see either full text of item or the first page. Users should also be able to go directly to the high-resolution CRT terminal and find material on a specific subject or from a specific source without having to use a low resolution terminal first.

D 4. Whenever possible, sequential commands should be manipulated by function keys, e.g., the user calls up the bibliographic record and can display images for the document by using a function key rather than entering image numbers or commands.

M 5. Commands must be compatible with other retrieval commands used in the online catalog.

M 6. All links between the bibliographic database and the images should be bidirectional, so that a user can move from the image to its bibliographic record as well as from the bibliographic record to its images.

D 7. Once the user has retrieved the set of images that represents a document a user will be able to directly access the table of contents, index, body of document, or any other sections of the document. These sections will be marked to the system at time of input. Thus the user will be able to proceed directly to the first page of these sections without browsing through the document.

D C. Access Time

Access time to the first image in a document from terminals connected directly to the optical disk system shall be a maximum of 9 seconds. Browse time to subsequent images shall be no more than 2 seconds.

D. Display at terminals (unless otherwise indicated these requirements apply to optical disk workstations with special terminals)

Terminal displays should provide:

D 1. Windowing, whereby two pages can be displayed simultaneously, or a close-up of parts of a page (see also E.2 below) can be displayed along with a full page.

M 2. Sufficiently high screen resolution to be legible for browsing over a period of hours, as well as comfortable to read. The user should be able to read fine text of size 4 point, hand-writing, pencil lines, and marginalia.

D 3. Color

D 4. Ability to suppress display of certain images to certain users.

M 5. A display at a minimum resolution of 300 pixels per inch (PPI) by 150 PPI, although higher PPI is desirable, if found to be economically feasible.

D 6. A display at the resolution in which the image is stored.

E. Search and display techniques

The following capabilities are needed:

M 1. The ability to browse rapidly in a logically sequential manner through a document at variable speeds. Browsing must include skipping forward and backward by N images, when N is a number entered via the keyboard.

D 2. The types of windowing whereby a patron could maintain either the index or table of contents to a book in a window while browsing through pages of images and text from the book, or could maintain the bibliographic record in a window while browsing through the pages of a book or document.

D 3. A vertical image should be able to be rotated 90 degrees to the horizontal without loss of resolution.

F. Input

1. Scanning

M a. If there is a bibliographic record for the document, the system will automatically link images to be scanned with the bibliographic record either through direct entry of the record number or through scanning a barcode. All scanned images are sequentially numbered by the system, and linked to the bibliographic record number for subsequent retrieval.

M b. In all cases, the system must maintain an image that is not anamorphically distorted, e.g. pages must not be scanned at 200 PPI on the vertical side and at 400 PPI on the horizontal side.

M c. All input will be at a minimum of 300 PPI (although 400 PPI is highly desirable, if found to be economically feasible.) The system will use a two dimensional isomorphic compression. That is, the compression algorithm must allow for the exact reconstruction of original digital image when the data is decompressed for reviewing.

D d. As part of the scanning operation, the scanner should have the option of setting the scan to produce halftone of continuous tone areas using a pseudo-halftone technique.

D e. Innovative scanning techniques, such as dynamic thresholding and generation of pseudo-halftone patterns are desirable (enhancing the density of a local area as opposed to the entire document).

M f. The operator should be able to vary the threshold either for a single image or a single session of scanning.

M g. The operator should be able to adjust for poor legibility or contrast problems by manipulating a lightness/darkness setting similar to the lightness/darkness setting on an office copier.

M h. If several images were required for a single page due to varying the settings of lightness/darkness (threshold levels) and pseudo halftone settings, all copies of the page should be logically linked together by the system.

M i. The system must allow for validation of the scanned document immediately after scanning. If inspection shows the document must be rescanned then or later, the rescanned image will logically replace the originally scanned image.

M j. System must allow images to be resequenced prior to writing to optical disk.

M k. System must allow for images to be inserted in sequence before writing to optical disk.

D l. The scanner will read single pages. If pages are double-sided, the scanner will read both sides simultaneously. The scanner will have the ability to read bound volumes without removing the binding, e.g. using a cradle reader.

M m. The scanner will have the capability for automatic page feeding and will allow for the manual feeding of documents when necessary.

M n. The scanner must accommodate documents up to 11 x 17".

M o. The system must allow for the capture of black and white documents, color as monochrome images, and any combination of text, continuous tone, half-tone, line art, line drawings on poor quality paper, type on low contrast paper, in short anything that could be found in a typical commercial catalog.

D p. If possible the system should capture, store, and display color as color.

M q. Scan time per image should not exceed 25 images per minute at 300 PPI (although 400 PPI is highly desirable).

D r. The system must be able to scan microfilm, or other microform images.

D s. The scanning system should not preclude the later addition of color laser copier technology, which is expected to be available within the next two years.

2. Keyboard data entry

M a. Input staff can link to image data such things as brief bibliographic record, document control number, document bar code, etc.

D b. Input staff can identify on the system any sections of the document, such as table of contents, index, bibliography etc. and provide display text naming each identified section. Later, on retrieving the document, users will be able to directly access the named sections of the document without having to browse through the entire work.

M c. Links between manually entered data and scanned data can be made at any time and by a variety of methods. This is to cover the times when a document has been input but the MARC record for it is discovered/entered later. It will be necessary to create the link between the images and the MARC record subsequent to the input of the image. This should be as easy as doing the link at the time of inputting the image.

M d. At no time will the "write once" nature of optical disks preclude the addition of links to or from the images stored on the disks. Links will be made to and from the document number, page number, or image number within the recorded document at scan time.

G. Storage

D 1. When it is determined that an image, page, or document needs to be rescanned, the system will provide a capability for integrating newly rescanned images with the images that they are replacing and establishing all necessary links with the new images that were in place for the older images.

M 2. The system will have the ability to store documents online temporarily on a magnetic medium by document control number and scanned image number for examination online of quality of image and/or tagging information.

M 3. There should be an option allowing for the optical data disk not to be closed at the end of a write session so that it may be written on until full.

M 4. The system should support retrieval from multiple disks, such as, but not limited to, the jukebox concept. If there are better systems for storage of multi-packaged information containers, this would be desirable.

M 5. The system must provide for automatic update of the bibliographic record to indicate that the text of the document is available on optical disk.

H. Output

1. To Printer

M a. Resolution must be sufficient to reproduce images from black and white documents, color as monochrome images, and any combination of text, continuous tone, half-tone, line art, line drawings on poor quality paper, type on low contrast paper, in short anything that could be found in a typical commercial catalog.

M b. The centralized offline batch printer should operate in a normal office environment, and must be able to print a minimum of 25 pages per minute of images at 300 PPI and ASCII text on 8 1/2 x 11" or 11 x 17" sheets.

D c. The centralized offline batch printer should be able to print at the same PPI as the scanning system allows.

M d. The work station printer must operate in a normal office environment and be able to print 5 pages per minute at 200 PPI, of 8 1/2 x 11" text. If feasible and economical 300 or higher ppi is desirable.

M e. The user must be able to print the page on current display as well as to select pages for printing based on boolean and other search modalities, such as a grouped span of images, specified number of images etc.

M f. It must be possible to print from temporary (e.g. magnetic) buffer storage as well as directly from optical disk.

M g. The user must be able to request printout based on document numbers or image numbers, bypassing terminal display.

M h. It should be possible to adjust speed in the batch printer to maximize quality.

M i. The system should have the capability of counting copies made by batch and work station printer and applying a "charge back" routine to allow for reimbursement of costs (much as an auditron governs photocopying).

D j. The batch printer must be capable of double-sided printing.

D k. The batch printer should be able to print in color.

M l. The system must have the ability to monitor orders for the batch printing system and to report the print load so policy decisions concerning usage, cost etc. may be implemented. It must allow for restrictions on types of printing to be applied, e.g., only x number of pages to be printed from a printer in a given time frame by a given user.

D m. The system must have the ability to print the bibliographic record for the document and a copyright notice on cover sheet or individual pages.

M n. It must be possible to stop the printing at a work station or batch printer at any time.

3. To other media

M a. The system must have the ability to output CCITT group 3 and 4 telefacsimile.

D b. System should be able to output images to COM

D c. The system must be able to direct output through various communications media, e.g. telephone, local area networks, satellite, etc.

D I. Character recognition and conversion of scanned images to ASCII characters.

When all the data on the document is in text and the content rather than the form of the information is primary, it is desirable to be able automatically to convert the text (or selected portions of the document such as tables of contents, indexes, etc.) to ASCII characters using character recognition software.

D 1. Character recognition software should be able to work on images already stored on optical disk to convert them to ASCII characters.

D 2. If scanned data is stored as ASCII characters, all indexing software used in the online catalog should be usable on the scanned data, e.g. author, subject, title, keyword.

D 3. The ASCII text should be easily integrated into the online catalog displays of bibliographic records as well as the displays of scanned images.

J. Maintenance

For ease in maintaining the optical disk system, the following are required:

M 1. Ability to track location, status of individual disks

M 2. The optical data disk must provide a guaranteed bit error rate (BER) of 10**12 after ten years of storage. System software must support recopying of disks before the end of their useful lives.

M 3. System operation reports should be available on demand in printed form and as a screen display.

M 4. All equipment should operate in a normal office environment. All media should withstand storage and use in a normal office environment, and transport via well-packaged mail.

M 5. The system should have the capability to provide information about specific disks, e.g., % of disk available for new images, % of disk occupied by error correction etc.

VII. Other Vendor Responsibilities

M 1. Provides and maintains hardware. Will need to agree to a maintenance agreement where response time to hardware problems are specified e.g. must have engineer on site within x number of hours for any problems that make system use impossible.

M 2. Writes and maintains software; debugging should be part of the agreement.

M 3. Insures compatibility with existing online catalog installation.

M 4. Provides test data on archival life of proposed storage medium so its use in preservation can be accurately estimated. Provides periodic updates of reports of archival life of storage medium based on continued testing and actual use experience.

M 5. Provides complete system instruction and description, including up-to-date manuals available when system is delivered, data dictionary if appropriate, training classes, operator manuals, etc.

M 6. Supports and maintains schedule of implementation for the life of the contract; provides monthly written status reports.

M 7. Insures that new optical disks are available as needed on schedule.

M 8. Follows and implements all national standards as they develop in this technology including storage, communications, and interchangeability between disks. As storage standards change and are implemented by the vendor, vendor insures compatibility with previously stored disks and software. If it is necessary to convert older disks to a newer format, vendor insures that software and hardware is available to perform conversion.

Proposals

Kent State University Libraries

REGIONAL CAMPUS LIBRARIANS

DEVELOPMENT OF A REGIONAL CAMPUS LIBRARIES' DATABASE

A Proposal

Regional Campus Libraries Automation Subcommittee

Jeanne Langendorfer
Pat Murphy
Jay Molendyke

May 27, 1986

REGIONAL CAMPUS LIBRARIES

DEVELOPMENT OF A REGIONAL CAMPUS LIBRARIES' DATABASE

A Proposal

INTRODUCTION

The Kent State Board of Trustees approved the purchase of Northwestern University's NOTIS (Northwestern Online Totally Integrated System) system as the University's integrated library system at its December 1985 meeting. The NOTIS software has been installed on the University's IBM computer and the Kent Campus Library anticipates implementing the NOTIS system in 1986/87.

The Regional Campus Libraries must be a part of the University's integrated library system to allow them to provide the information services students and faculty are beginning to expect and demand in the 1980s. Benefits to the Regional Campuses would be significant. The Regional Campus Libraries would increase circulation and, in fact, overall utilization would increase. In addition, more effective interlibrary cooperation and greater access to the materials at the eight University libraries would be achieved at minimal cost. This is especially valuable as the book budgets at the Regional Campus Libraries have shown little or no growth in recent years.

When the Kent Campus has an operational online catalog next fall, the Regional Campus Libraries will have the potential for direct dial-access to the system. Dial-access is dependent on the use of the system by others, involves relatively expensive long distance charges, and requires additional equipment. Direct access is preferable and may be accomplished through the proposed online network which is currently being considered for linking the seven Regional Campuses with the Kent Campus.

Regardless of the type of online access which becomes available to the Kent Campus NOTIS database, the holdings of the seven Regional Campus Libraries will not be included in the database unless immediate action is taken to convert these printed card catalogs into an appropriate machine-readable format.

The process of converting a library's holdings, as represented in the card catalog, into a machine-readable format is know as retrospective conversion. The Regional Campus libraries have approximately 300,000 records/volumes which have to be "retroconverted." A retroconversion project must be undertaken so these records can be integrated into the NOTIS system.

It will take approximately 2-3 years to retroconvert all of the 300,000 records in the Regional Campus Libraries. However, these records can be added to the NOTIS database as they are converted on a monthly or weekly basis. These records will then be accessible as they are produced.

Once this database is produced and once the Kent Campus has an operational online catalog it will be necessary to determine the type of access (e.g., dial-access or through an online network) and the type of terminals (e.g., microcomputers or dumb terminals) which are required. Neither the type of access nor the type of hardware required can be accurately identified at this time. However, it is clear that a necessary first step is to build a Regional Campus Libraries' database.

The Regional Campus Libraries are requesting funding for a project to convert the 300,000 bibliographic records from their seven libraries into a machine-readable Regional Campus Libraries database. This database will then be loaded into the Kent Campus NOTIS integrated library system.

This paper outlines the rationale for an immediate retroconversion project for the Regional Campus Libraries and details the various options and their costs.

RATIONALE FOR A RETROCONVERSION PROJECT FOR THE REGIONAL CAMPUS LIBRARIES

The following represent the basic rationale for initiating a retroconversion project for the Regional Campus Libraries and for integrating the Regional Campus Libraries' holdings into the University's NOTIS system:

(1) The Kent Campus Library has received Board of Trustee approval to purchase the NOTIS integrated library system. Given the fact that the Board of Trustees has made the automation of the Kent Campus Library official policy and has committed substantial funds to support the implementation of NOTIS as the University's integrated library system, the Regional Campus Libraries should be

included in this plan. The retroconversion project is the essential first step in this process.

(2) Regional Campus students and faculty need direct access to the bibliographic records (e.g., information in the card catalogs) in both the Kent Campus Library and the other Regional Campus Libraries. Students and faculty frequently attend and/or use more than one facility. Similar services and access to the system's collections should be provided at all institutions serving Kent State users.

(3) Kent Campus students and faculty could benefit from direct access to the bibliographic records of the Regional Campus libraries. All of the Regional Campus Libraries have some materials which are not available in the Kent Campus Library. All Kent State students should have access to the materials purchased for these eight libraries.

(4) Students entering Kent State have been increasingly exposed to new technology in their high schools and public libraries. The Regional Campuses, as representatives of Kent State University, should not be perceived as out-of-date facilities. Students now expect automation and technology in the classroom and in the library.

(5) The Regional Campuses are an integral part of Kent State University and represent the University in their local communities. Direct access to the collections of the Kent and other Regional Campus Libraries will enable the Regional Campuses to better serve their patrons.

(6) The Regional Campuses and their libraries were viewed once as autonomous entities which sought to provide higher education for those unable to attend the Kent Campus or for those who did not want to pursue a baccalaureate degree. New courses and new degree programs have been developed which challenge the capabilities of Regional Campus Libraries to meet the student and faculty needs with their limited resources. Minimal collection development budgets have increasingly forced Regional Campus Libraries to rely on each other and the Kent Campus library.

(7) Any proposed network linking the Regional Campuses with the Kent Campus will benefit library users only if advance work is undertaken to convert the Regional Campus Libraries' holdings into a machine-readable format. Student and faculty access to an online library catalog involving all eight Kent State campuses could support the creation of such an electronic network.

(8) The Regional Campus Libraries believe the benefits of increased library utilization by both students and faculty and increased availability of the materials contained in the University's eight libraries justify the costs for including the Regional Campus Libraries as part of the University's integrated library system.

RETROCONVERSION OPTIONS & COSTS

Three methods of retroconversion are currently available:

(1) Online Bibliographic Networks
This involves direct telecommunications with a major online bibliographic network. It requires membership fees, equipment costs, telecommunications charges, and substantial per use fees. OCLC is used as an example.

(2) Microcomputer-Based Interface with Online Networks
This is a relatively new intermediate possibility which allows a library to use microcomputers to prepare lists of items to be processed against the files of a large online network. No telecommunications charges are involved, but a per use fee is assessed. The OCLC/MICROCON system is an example.

(3) CD-ROM Technology
This new technology represents a significant departure from other methods of retroconversion. This subscription service utilizes a self-contained CD-ROM unit and is available at the library undertaking the local retroconversion project. No telecommunications charges are required and, significantly, no per use fee is involved. An example of a CD-ROM retroconversion system is Bibliofile. The Bibliofile system uses the new compact laser disk (CD-ROM) technology to store more than 3,000,000 bibliographic records on four compact laser disks.

Table I compares these three methods of retroconversion. The following assumptions were made in making these projections:

(1) 300,000 bibliographic records will require retroconversion in this project.

(2) Two terminals/microcomputers will be used.

(3) Approximately 2-3 years will be required for the project, regardless of system used. Equipment and communications costs are given for three years.

(4) Approximately 80 hours/week will be required to accomplish the project in 2-3 years using two terminals.

(5) Staff requirements and costs are not included for any of the systems. Each system would require similar amounts of staffing. OCLC/MICROCON and Bibliofile could be accomplished with minimally trained student or part-time workers. OCLC would require slightly more experienced staff.

(6) Staff costs would total approximately $15,000/year based on 80 hours/week for 50 weeks/year at #3.35/hour (80 x 50 x $3.35 = $13,400). Supervision would be provided by the library housing the equipment. Student workers could be employed for this project.

Appendix A includes the results of the test of the Bibliofile system by the Regional Campus Libraries.

TABLE I

SYSTEM	Equipment Costs	Communications Costs	Per Use Charges	Total Per Use Charges	TOTAL COST	"Hit Rate"
Bibliofile	$ 8,820[a]	0	0	0	$ 8,820	78%[b]
OCLC	$12,960[c]	$ 9,000[c]	$ 1.25[d]	$375,000	$393,900	98%[b]
OCLC/MICROCON	0[e]	0	$ .40	$120,000	$120,000	91%[b]

[a] Until July 1, 1986, any library subscribing to Bibliofile for a three year period (total advance payment of $4,410) will be given the following equipment:

IBM PC-XT-compatible computer with 512K memory and 10 megabyte hard disk; and
International Standard Hitachi 1502s CD-ROM drive

This equipment becomes the property of the purchasing library.

[b] Hit rate refers to the number of records actually found on the system. Records which are not matched will have to be manually input. Both Bibliofile and OCLC allow for such manual entry of records.

[c] OCLC Terminal Costs:
$180/month x 2 terminals x 36 months = $12,960
Telecommunications:
$250/month x 36 months = $ 9,000
NOTE: Other miscellaneous OCLC costs (e.g., $480/year OCLC/OHIONET membership fee) are not included.

[d] Based on Non-prime time access. Prime time access is $1.40.

[e] OCLC loans M300 PC workstations for use in the project.

DECISION CONSIDERATIONS

Two major factors must be considered in selecting which retroconversion option to use for the Regional Campus Libraries:

(1) Which is the most cost-effective?; and

(2) Does the most cost-effective system produce an acceptable product?

(1) Cost-Effectiveness

Bibliofile:	$ 8,820 + staffing
OCLC:	$393,900 + staffing
OCLC/MICROCON:	$120,000 + staffing

Bibliofile is clearly a more cost-effective option.

(2) Product

Bibliofile:	78% Hit Rate (Based on RC test)
OCLC:	98% Hit Rate
OCLC/MICROCON:	91% Hit Rate (Based on OCLC predictions)

These figures indicate that approximately 80% of the materials owned by the Regional Campus Libraries will be found in Bibliofile. Bibliofile has the capability to allow manual input of any records not found into the Regional Campus Libraries' database. This process simply involves retyping the catalog card onto a formatted Bibliofile screen.

In the test conducted by the Regional Campus Libraries, 65% of the items not found on Bibliofile were published prior to 1965. Since these materials are older and more ephemeral, action on these could be postponed until the end of the retroconversion project (see Appendix A).

Conclusion: Bibliofile will produce very acceptable results for a very competitive and cost-effective price.

Kent State University Libraries

PROPOSED BIBLIOFILE PROJECT SCHEDULE

(1) Project Schedule

1986-87

System Subscription Cost:[a]	$ 8,820
Staff Cost:[b]	$15,000
Production:[c]	120,000 records converted

1987-88

System Subscription Cost:[a]	$ 0
Staff Costs:[b]	$15,000
Production:[c]	120,000 (Total=240,000)

1988-89

System Subscription Cost:[a]	$ 0
Staff Cost:[b]	$ 7,500 (1 FTE)
Production:[c]	60,000 (Total=300,000)

TOTAL 3-YEAR PROJECT COST: $46,320

[a] For the advanced payment of a three-year subscription ($4,410/unit) before July 1, 1986, Bibliofile will provide a IBM PC-XT-compatible microcomputer and a CD-ROM reader. This hardware becomes the property of the purchaser.

[b] 80 hours/week @ $3.35 x 50 weeks = $13,400/year
$15,000 allows for some overlap, unexpected staff changes, and training.

[c] 30 records/hour x 80 hours/week x 50 weeks = 120,000 records/year
The test of Bibliofile conducted by the Regional Campus Libraries indicated that a rate of 1 record every 2 minutes was conservative.

RECOMMENDATION

It is our recommendation that two subscriptions to the Bibliofile sytem be funded for the Regional Campus Libraries for the purpose of retroconversion. This will provide two workstations and a three year period for completing the project.

Total cost for the subscriptions, if committed to before July 1, 1986, will be $8,820. Staff costs of approximately $15,000 per year are also requested. The projected three-year cost of the total project is $ 46,320.

After receiving approval to subscribe to the Bibliofile system, a project plan will be finalized which specifies exact procedures, locations, and other technical details for completing the project as outlined in this paper. The use of Bibliofile, as proposed in this paper, will not directly affect any other procedures currently employed by individual Regional Campus libraries.

APPENDIX A

In February 1986, the Regional Campus Libraries' holdings were tested using Bibliofile. Each campus contributed 100 shelflist cards selected randomly. These titles were searched in Bibliofile. Listed below are the number of records located (hit rate) for each of the campuses:

Campus	Hit Rate %
Ashtabula	70
East Liverpool	84
Geauga	82
Salem	78
Stark	80
Trumbull	82
Tuscarawas	76

The average hit rate was 78.85%. Of the items not found, 65% were older than 1965.

A second 100 shelflist cards from Geauga were selected and tested. The hit rate was 76%. The records were searched, located, modified, and saved in 2.25 hours. The records not found were searched using several different strategies in this same 2.25 hours. This is an average of 1.35 minutes/record tested or approximately 45 records/hour.

Massachusetts Institute of Technology

A PROPOSAL TO PROJECT ATHENA
FOR
A VIDEODISC RESOURCE SYSTEM

Abstract: A study is proposed to develop a visual resource system using videodisc-stored images linked to a searchable database in Rotch Visual Collections as a means for delivery of visual materials to the studio and the design workstation in the School of Architecture and Planning.

This proposal is for an initial study to develop and test a visual resource system which would serve as a model to provide access to visual collections for study and teaching in the School of Architecture and Planning. The goal is to make available the library's visual holdings in the studio/classroom and to provide a visual research tool as a resource for the student at the integrated Athena design workstation. The pilot study proposes to do this by storing on a videodisc selected images from the Rotch Visual Collections and linking it to a database comprised of the catalogue records developed and maintained in the library. As an initial step it would test student use in an installation in the library and explore the delivery of images to the School's Computer Resource Laboratory. It will look toward later extension of the capability of delivering images to other parts of the School which are housed in N52, W31 and E15.

Rotch holds a rich store of some 7,000 slides and photographs of Boston, materials which are heavily used by the students and faculty of the School for a variety of courses in architecture, urban studies and planning. While demand is high for these images, manual access is slow and random. Frequently images found only in one category may illustrate a variety of issues for different users. For example, slides of the waterfront or Quincy Market area are used to discuss commercial redevelopment, housing, urban change, transportation, pedestrian access, public space, and zoning to name only a few topics. Patient search may net good results for the more persistent, but often excellent materials are overlooked by the rushed or the less experienced, particularly the undergraduate student unfamiliar with the collections or their arrangement. A computer-linked videodisc would allow review of the whole collection with ease and permit searching for specific subjects such as tri-decker housing in Dorchester or aerial~~s~~ views of the Dudley Station area. Search and choice would not be limited to the library. Linked to the workstation in the Computer Resource Laboratory, the entire range of images could be delivered to terminals remote from the library at any time, and for any duration of time. If the disc-stored images were coupled with a graphics capability to permit overlay, implications of various design decisions could be studied in context. For example, at the workstation, changes could be plotted directly onto actual views of a site, or proposed housing inserted into a neighborhood.

Related Efforts

Under the umbrella proposal developed by Ferreira, Miller and Purcell "A Graphics Environment for Educating Architects and Planners" The School of Architecture and Planning has under development several computer projects which will form a design support system. These include specific projects in building technology, environmental behavior, site planning and design modelling. The videodisc resource project aims to broaden that capability by placing a searchable body of primary visual material directly into the design environment; to add a new visual dimension to those other Athena funded projects.

At the same time, as a library project, it would look at a means of delivery and access to a unique MIT collection with special relevance to the students of the School. The MIT Libraries are submitting to Athena a proposal to link information services and an online catalogue of the Libraries holdings to the Athena network. Over a four-year period of development, campus-wide access to bibliographic records of the Libraries collections will be made possible. Visual materials, however, are an exception. These have traditionally been catalogued separately, using a specially-designed system. The Rotch Visual Collections plans to create a separate online catalogue, with planned tie-in to the Libraries online catalogue in 1989. Work undertaken in the feasibility study will form the basis for developing a sensitive and appropriate method for access to the Visual Collections by its major constituency and will complement the Libraries initiative. It could also serve as a prototype for handling other types of visual materials through the Athena system.

The Images of Islamic Architecture Videodisc Project

This proposal is an outgrowth of a related videodisc project of The Aga Khan Program for Islamic Architecture. In that instance, a videodisc of over 30,000 images of Islamic architecture is controlled by a computerized database mounted on the IBM mainframe using the Focus database management system. The work proposed here draws on the experience gained by Smith, who is Director of the Videodisc Project, and by the staff in indexing records for the database and in the creation of the disc-database interface to apply methods to develop a system for delivery of images to the School.

Both that and the proposed work draw on the experience of the Architecture Machine Group and Professor Purcell in creating the Visual Information System which holds an earlier group of slides from the Visual Collections and lays out a method for records display in a system called Archfile.

METHOD

The schedule for the project allows for compiliation of materials during the summer 1985, development of the database and requisite software during the autumn '85 and early spring'86 and to implementation and testing of the system's installations in spring term '86.

1. Materials on Boston will be selected from the teaching collections the Kepes/Lynch photograph collections, the materials compiled for the Copley Square Competition and the Boston Globe Conference in 1984. Other materials as appropriate, or other collections could be added. Summer '85.
2. Selected slides and photographs will be transferred to videotape or filmed for the videodisc during the summer '85 when they will not be needed for classes. The videodisc will be mastered and available by early fall '85.

3. Using the database management system Informix, slide cataloging information will be entered into a computer in Rotch Visual Collections beginning summer '85 and extending through spring '86. Some justification of records and additional subject indexing will be necessary.
4. Computer-disc interface and other software will be written in C, autumn '85. Initial work, in Rotch Visual Collections.
5. Installation of the disc and link to design work station in the Computer Resource Laboratory, spring '86. Test use in Rotch Visual Collections and the above location; others as possible.

BIBLIO: A Library-based Academic's Workbench

Richard K. Belew
Maurita Peterson Holland

12 May 1986

Abstract

This proposal recommends the development of a suite of software systems — dubbed collectively as BIBLIO — that offers to connect members of the University community to the University Library, in ways that have not been possible without the computers and communcations facilities only now becoming available. We begin by sketching the wide range of informational resources becoming available to University members, and then mention some of the work we have done previously on the Information Exchange Facility (IEF) project towards making these resources more available by computer. The bulk of this proposal is a description of the four levels of the BIBLIO system, corresponding to the special information needs of: individuals, "interest groups", University-wide use, and use outside the University. We are proposing that BIBLIO be implemented in four stages corresponding to these levels, beginning with a version designed for the individual and moving towards a system connecting to information resources outside this University. We then present our view of what the Library of the future might be like. From the this perspective, the goal of the BIBLIO project is to suffuse librarianship throughout the entire text preparation and retrieval process. We see BIBLIO as a set of mechanisms that help librarians generate indexing structures for materials not traditionally part of the library, and mechanisms that extend indexing structures built by librarians to people who would otherwise be unaware of them. We conclude with a two year time-table and budget that will lead to the implementation of the BIBLIO system.

Table of Contents

1 Perceived needs 3
2 Previous Work: The IEF Project 5
3 BIBLIO 5

  3.1 Level One: Personal information collections 7

    3.1.1 System Characteristics 7

  3.2 Level Two: Interest groups' information resources 8

    3.2.1 System characteristics 8

  3.3 Level Three: University Information resources 9

    3.3.1 Brokering of University information resources 10
    3.3.2 Shared indexing 10
    3.3.3 System characteristics 11

  3.4 Level Four: External information resources 11

4 An on-going role for the Library 12
5 Implementation strategy 13

  5.1 Time Table 13
  5.2 Project administration and personnel 14

6 Budget 15

BIBLIO: A Library-based Academic's Workbench

Richard K. Belew
Maurita Peterson Holland

12 May 1986

1 Perceived needs

Traditionally scholarly knowledge and research information has been organized, accessed and housed within the walls of the University Libraries. The formidable task of keeping abreast of knowledge output in the broad range of University interests or potential interests, acquisition of those materials and provision of bibligraphic access to the collections consumed substantial staff and resources. The Library was viewed foremost as the repository of accumulated wisdom and creative thought.

During the past several years, the Library has also changed dramatically in its ability to provide bibliographic access to printed information.

- online access to shared catalogs of other university libraries (RLG/RLIN) provide information about resources around the country, not just those available in the local library
- through partial subsidy, the Library offers online databases from vendors (Dialog, BRS, SDC, etc.) providing indexing and abstracting of titles, many of which were not held locally and many of which are available for purchase online
- even the considerable collections budget which the University has consistently provided the Library is not sufficient to purchase and bring to the Library shelves all of the most important scholarly output in many fields.

It is not surprising, therefore, that several factors have caused a change in the "single source" concept of information distribution and use on the Campus. Among them are:

- the physical size of the campus and growing interdisciplinary nature of both publications and researchers has fostered the growth of departmental or laboratory collections which serve as close at hand working libraries. In a survey done 10 years ago, 80 such collections were located and a suspected additional 80 were identified.
- distributed computing now available to end-users puts the user in direct contact with the same online information sources which are available to the Library
- the University Library's circulation system on our Campus provides a limited database of University Library's holdings
- document delivery services now offered by the University Library provide information to the user's office
- the user is now being forced to decide how grant funds, departmental accounts and personal money should be spent acquiring information.

The microcomputer, together with telecommunications capability put the individual faculty/student in the center of the information universe. In fact, the BIBLIO project seems a perfect example of the kind of "content innovation" that is made possible by the "systems innovation" being provided by CITI and the Computing Center[1].

For the first time University members can build unique files of materials in their personal collections, locate materials shared by others in their department, lab, research group, network, etc., access University Library holdings and the databases of nationally-vendored services. The key insight of the *near—library* concept is that: individuals first use the collections which are nearest, easiest and quickest to access, and only move on in ever-broadening searches in terms of distance, time and cost if their information need has not been satisfied.

This shift of access capability from the Library to the individual offers an exciting and unrealized potential for the Library. Certainly the Library's role need not diminish in the scholarly information search. Rather it needs to play a central role in developing a user-based system only possible when it is provided to the campus community. In its new role, the Library provides training to users, staffing/procedures/guidance in organizing and maintaining departmental collections and maintenance/administration of the databases. In fact, such a development is in the best interest of the University and therefore its scholarly community because it provides bibliographic access to individual collections, fosters sharing of information and could be used as a vehicle for gaining access to materials otherwise unknown on the Campus since they reside outside the bibliographic control of the Library.

Almost 2 years ago, the Engineering Libraries began to build a prototype of such a system, funded by the College of Engineering with help from the University Library. In the course of evaluating existing software, we discovered that:

- no existing software meets the needs of the scholarly community
- software required to meet most of the objectives is highly expensive and not generally available for all of the University's machine environment (IBM-PC, Macintosh, Apollo, mainframe)
- there is wide-spread interest in improving bibliographic control/access on the Campus.

Many of the problems addressed by a system like BIBLIO are due to the special nature of free text as a type of information. In particular, while it is possible to completely ennumerate all the "facts" contained in a database record, this is impossible with free text. Information retrieval systems cope by selecting the most salient characteristics of a piece of text and "indexing" it under these characteristic "keywords." As we move towards the design of a truly integrated Academics' Workbench, a concern with the management of textual information becomes central. We believe many of the same techniques we are exploring with the BIBLIO facility will prove helpful with the management of other, non-bibliographic text as well. For example, a version of the adaptive, shared indexing procedure outlined below could be used to manage text that has not (yet) been published. Even text that is in the process of being created could benefit from BIBLIO. There are, for example, significant similarities between the outlines developed by "pre-writing" tools (e.g., ThinkTank) and the extensible indexing taxonomies we manage. The major differences are in the of level of refinement and degree of idiosyncracy. Both of these issues are addressed by the BIBLIO project.

---

[1] See "A user perspective of computer workstation integration", D.E. Atkins and R.L. Phillips

This work has convinced us that the University needs a bibliographic facility — which we have come to call BIBLIO — which is available to students and faculty at low or no cost, meets academic needs, runs on all machines in the Campus environment, exploits the rich collections which exist outside the formal Library system and offers the potential to produce "knowledge products" which represent the academic output of our scholarly community.

2 Previous Work: The IEF Project

For approximately two years now, the Information Exchange Facility (IEF) Project has been actively exploring a new role for personal computers in the University Library setting. Funded as a joint project between the Engineering College and the University Library system, the IEF Project began by establishing personal bibliographic collections on IBM/PC's for two "targeted" groups of faculty in the Computer Science and Chemical Engineering departments. After evaluating a number of available software packages, a number of copies of a program called REF-11 were purchased for faculty in these two groups.

For the group in computer science, the emphasis was on capturing citations these faculty members knew maintained manually. One CSE faculty member (Prof. Jain) had a large bibliographic collection in a machine readable file, and we brought this across into REF-11 for him. In total, we have captured approximately 2,500 bibliographic citations for this group in CSE.

In Chemical Engineering, the emphasis has been on supporting faculty access to the Chemical Abstracts Online Service (CAS). We have written a program which takes a CAS session log file and extracts the bibliographic citations and places them in REF-11. Also, we have begun to explore the possibility of processing existing word processing files for their bibliographic citations.

More recently, the focus of the IEF project has moved to a facility through which individual's personal bibliographic collections could be shared. We now have an elaborate bibliographic database defined, using SPIRES database software on MTS. This definition follows along the lines of the ANSI standard for bibliographic information. We have also developed links between the REF-11 databases and the SPIRES database on MTS which allow individual bibliographic collections to be uploaded to MTS. A facility for downloading information from SPIRES into REF-11 is under development.

Finally, we have been at work connecting these database systems to the word processing software used to actually format the bibliographic citations. We are now at work on a link between SPIRES and BibTeX, a utility working with the TeX Package. When finished, this link will allow users to export bibliographic information from our SPIRES database into BibTeX, allowing the full flexibility of text formatting.

3.1 Level One: Personal information collections

The goal of the Level One component of the BIBLIO system, which we will call BIBLIO-1, is management of a local bibliographic database. In many ways, this facility will be similar to traditional database management systems (DBMS) now available on micros. Our experience of the last few years, however, has convinced us that special characteristics of bibliographic data make standard DBMSs inappropriate for our purposes.

We invision BIBLIO-1 dealing with two basic types of information. First, it will manage bibliographic citations: titles, authors, publication information, etc. These will be enhanced with keywords describing the work, as well as annotations about the work, quotations from the work, etc. The second type of information managed by BIBLIO-1 is unstructured free text. The annotations and quotations associated with citations are one example of this form of information, but we do not believe BIBLIO-1 should be restricted to these examples. We foresee BIBLIO-1 users maintaining large textual databases of things they have written, messages they sent or received, electronic bulletin board items, etc. The issues involved with the management of this free-text information are very similar to those required to manage bibliographic information. We want BIBLIO-1 to provide facilities (e.g., like a general purpose indexing/search engine) that make it useful in all of these contexts.

3.1.1 System Characteristics

There are three general design constraints that must be considered early in BIBLIO-1's development. First, BIBLIO-1 must be made to interact closely with various word processing systems being used by its users. This means word processing specific characteristics (such as formatting control codes) must be variable parameters made available to BIBLIO-1, and not "written into the code." Second, while a key motivation of the BIBLIO project is that each of the levels must be integrated, we believe that it is important for BIBLIO-1 to work without the rest of BIBLIO. Unless BIBLIO-1 users can be assured that personal collections they are building will migrate with them when they leave the University, the probability of BIBLIO's acceptance will be much lower.

Third, BIBLIO-1 must be made to work on the wide range of machines used by various groups within the Unviersity. This means working in environments like, (on the low end) an IBM PC with 256K Ram and floppies or a (thin) Mac, to an IBM PC/AT with 1 Meg of Ram and a 20 Meg hard disc or a Mac-Plus with a hard-disk, to (on the high end) an Apollo or Sun or IBM-RT workstation. Getting BIBLIO-1 to run on this wide a range of environments will be one of the early, large challenges of the project. We are confident, however, such a design is possible, with individual configurations being specified parametrically.

We are assuming a 1200 baud phone line connection as the basic communication link from BIBLIO-1 to other layers of the system. This is a key assumption, and one that should be re-evaluated as more and more of the University community becomes connected via faster communication links (e.g., UMNet, the optical fiber network). Similarly, BIBLIO-1 design will be dramatically affected by optical disc technology. An example of how this communication mass storage trade off appears in the BIBLIO context will be mentioned in the description of BIBLIO-2, below.

The basic functions of BIBLIO-1 include:

- Standard database facilities (full-screen input, report generation, updating);
- Specialized support for the special types of data used for bibliographic information: free-text (e.g., wild-card string matching), authors, dates, keywords;
- Support for ancillary information (e.g., annotations, quotations) typically attending citations;
- Utilities for connecting this database to the process of word-processing, such as processing a manuscript and replacing "pointers" with correct citations and automatically constructing the bibliography.

3.2 Level Two: Interest groups' information resources

The focus of BIBLIO-1 – the individual user – was clear. The focus of BIBLIO-2 requires a bit more explanation. Most members of the University community are associated with one, and possibly more, special purpose interest groups. Typically, these are smaller academic departments or research labs within the larger departments. However, any number of topics can serve as the focus for an interest group, as the proliferation of CONFER and *FORUM conferences on MTS demonstrates. The goal of BIBLIO-2 is to provide library-like support for these interest groups. By definition, members of an interest group share common interests. Information collected by one of their group is likely to be valuable to the rest of the group. We see BIBLIO-2 as a foundation for collaboration between members of the group. In fact, this sort of collaboration already occurs informally and BIBLIO-2 will only serve to promote it. Departmental libraries and technical report series now maintained by some interest groups are exactly the sorts of information we imagine BIBLIO-2 maintaining.

Another important example of information maintained by interest groups that could be well supported by BIBLIO-2 are *collections* scattered across the University. Our Campus possesses myriad collections quite beyond the books, journals and reports traditionally thought of as library collections. These collections, however, are frequently of considerable educational merit and very often are without organized access. Examples are the collections of software purchased for either personal or departmental/lab use, the Stearns Collection of Musical Instruments, slides, photographs, art works, videocassettes, films, etc.

3.2.1 System characteristics

The basic software required for BIBLIO-2 will not be that different from that in BIBLIO-1. The information management requirements in these two situations are quite similar. The two major differences are first one of scale, and second consequences of the role BIBLIO-2 will play as a communication link between individual members of the interest group and connecting the interest group to the University resources of level three.

As Figure 2 suggests, we imagine the interest group's BIBLIO-2 facility acting as the hub of a star formed with members of the group. BIBLIO-2, then, will act fundamentally as a "server" storing the *union* of information maintained by individual members. The degree of redundancy between information stored on the BIBLIO-2 host with that stored on individuals BIBLIO-1 facilities is the communication mass-storage tradeoff alluded to above. For example, it may be appropriate to use the BIBLIO-2 host as a "citation server," for storage of most of the citation information common across users, and maintain only the individual's keywords and annotation on their personal machines. This design makes sense if individuals typically have machines with very limited local storage and high speed communication links to the BIBLIO-2 host.

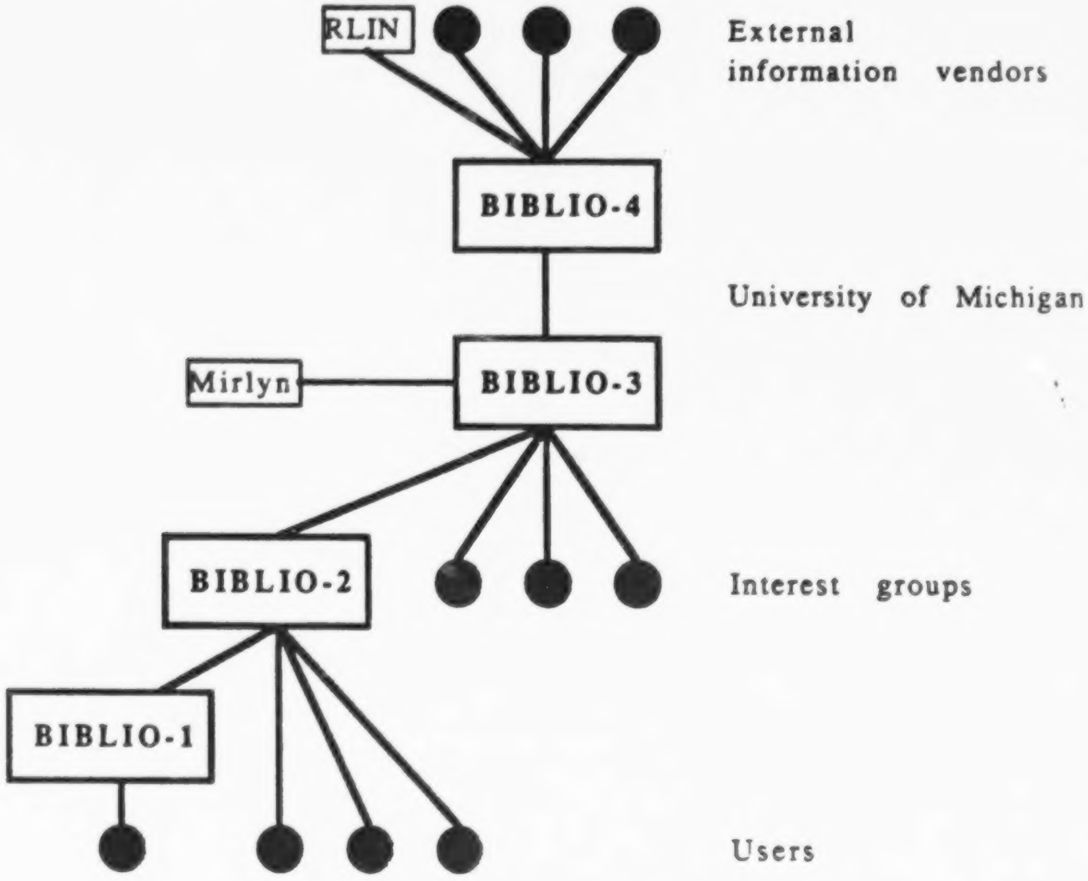

RLIN
External information vendors
BIBLIO-4
University of Michigan
Mirlyn
BIBLIO-3
BIBLIO-2
Interest groups
BIBLIO-1
Users

Figure 2 - A network view of BIBLIO

3.3 Level Three: University Information resources

Currently, the major University Library computer system is GEAC. This system is expected to be replaced with the MIRLYN system by January, 1988. MIRLYN is to provide online access to card-catalog information as well as availability (circulation) information for all works in the Library system. We therefore do not expect BIBLIO-3 to replicate any of these basic functions. However, we do intend to work on *linking* the MIRLYN system to BIBLIO-1 and BIBLIO-2, and to explore links between MIRLYN and the external information sources of Level Four.

5 Implementation strategy

5.1 Time Table

We foresee our work this summer as being divided between two tasks: a continuation of projects begun by the IEF project, and the beginning of the design for the BIBLIO System. We have a long list of tasks surrounding the work we have already done on the IEF project. If the BIBLIO project is not approved, these and a few other loose ends will occupy our group for the rest of the summer. However, if BIBLIO is approved, we intend to begin a design of the BIBLIO-1 facility this summer. We would like to have a top level, functional specification of BIBLIO-1 by Fall of 1986. Tasks for this summer therefore include:

- A careful review of existing software for strengths and weaknesses. A similar review was done two years ago and it can be used as a beginning, but a number of new and improved systems have become available since then and a review is necessary.

- A careful characterization of the computing environments we expect BIBLIO-1 to operate in. The approximate range of machines to be considered is mentioned above (Section 3.1.1).

- Layout of principle screen and report formats.

- Selection and procurement of a development hardware for use beginning in Fall 1986. We **are asking for an Apollo 3000 series machine (supporting Unix) as the main development** environment and to act as a prototype for full BIBLIO functionality, a large hard-disk to use for the databases we will be building, and an IBM-PC/AT and Mac-Plus for use in developing versions for these two machines. We also ask for the hardware and software required to network these three machines together, and to UMNet.

- Selection and procurement of database utility software. While there is much that is unique about the task BIBLIO is to solve, many of its most mundane functions are similar to standard database and information retrieval systems. We propose purchasing software that will keep us from having to "re-invent these wheels." SIRE, for example, is a very powerful and interesting text-retrieval system developed by academics and now marketed commercially. The source code for SIRE is available, and starting with it would let BIBLIO-1 development move very quickly. Low-level libraries (e.g., of file-access routines, user interface functions) are also available and may speed the development effort. We are also asking for development software. This may be cross-compilers from the Apollo onto the Intel 80286 and Motorola 68000, or it may be native compilers for the IBM-PC and Mac.

Beyond this summer our plans are less firm. We feel two person-years of development effort will be required for the development of the initial BIBLIO-1 facility. At the staffing level requested (viz., four CSE graduate student RA's at 50% each), BIBLIO-1 should be ready for Fall, 1987. We are proposing that the design of the additional capabilities required for BIBLIO-2 (largely communication links) be designed and implemented during the next academic year, to be ready for Fall, 1988. We are proposing that the design of the BIBLIO-3 system begin during this year, to coincide with the introduction of the MIRLYN system (scheduled to become operational in January, 1988).

THE KELLOGG PROJECT AT SYRACUSE UNIVERSITY

1986

EXECUTIVE SUMMARY

Syracuse University has an outstanding resource to share with adult educators throughout the world. For the past 37 years, we have collected a wide variety of print, visual, and media resources related to the education of adults, and today the Syracuse University Libraries house the most comprehensive assortment of English-language adult education resources in the world. The manuscripts alone total 650 linear feet of materials. These range from the complete records of past adult education organizations to selected papers of early and current leaders of the field. In addition, we have extensive circulating collections of books and audiovisual materials. In fact, two major professional organizations have designated Syracuse University an official archival repository for both adult education and educational gerontology. As a result, Syracuse now receives materials at an even greater pace than in the past.

Our Adult Education Collection has not yet been used to full advantage. Despite Syracuse University's commitment to adult education, we cannot keep up with the influx of materials; with our current human and financial resources, it will take us a number of years just to process what we already have, and even the already-processed contents of the shelves cannot be effectively disseminated beyond the Syracuse University campus. Therefore, only staff, students, and visiting scholars currently have access to them.

However, the adult education library-for-the-world is within our grasp. We as an institution have the ability to make this huge Collection available and useful as a common literature base for a widely scattered constituency. With our prominence in adult education, archives management, information studies, computer science, and other disciplines, we can set new precedents for information storage, retrieval, dissemination, and research, while furthering the now-urgent mission of adult education.

In view of the Kellogg Foundation's profound commitment to adult education and the Foundation's historic penchant for supporting significant innovations in that field, we are pleased to be able to propose this important project (called the Kellogg Project throughout) at Syracuse University, which we would like to name the Kellogg International Center for Adult Education Information Exchange.

We bring to the Project a tradition of multidisciplinary collaboration and an assortment of resources without parallel anywhere. Both the University and the surrounding community are noted for their strength in adult education. The School of Education at Syracuse University has a well-established graduate program in adult education, as well as expertise in instructional development, cultural foundations of education, distance education, and educational-resources dissemination. The Syracuse University Libraries have already-established systems for the electronic transmission of information, upon which we will build. We have experts on dissemination in our School of Information Studies, and faculty in the School of Computer and Information Science and in the Center for Advanced Technology in Computer Applications and Software Engineering (CASE Center) offer logic-programming tools developed on campus. Our University College staff will provide assistance with conference planning and developing an independent-study master's degree program. Finally, the Center for the Study of Citizenship within the prestigious Maxwell School of Citizenship and Public Affairs will bring together the finest thinkers at Syracuse University to study the Project in its societal context. The Project will draw upon all of these resources to create a coherent, unified, and far-ranging Center for Adult Education Information Exchange. (See Appendix A for more information about Syracuse University.)

The Project has three components (see Figure 1), each designed to tap the enormous potential of the Adult Education Collection. They are as follows:

- Dissemination based on optical-digital storage and computing technologies (including logic programming), concomitant professional research to make optimal use of the Collection materials, and continuing development of the Collection.
- Creation of an electronic network and software design for interactive access to information.
- Pursuit of educational opportunities, such as research, independent study, and conferences.

Figure 1: THE KELLOGG PROJECT

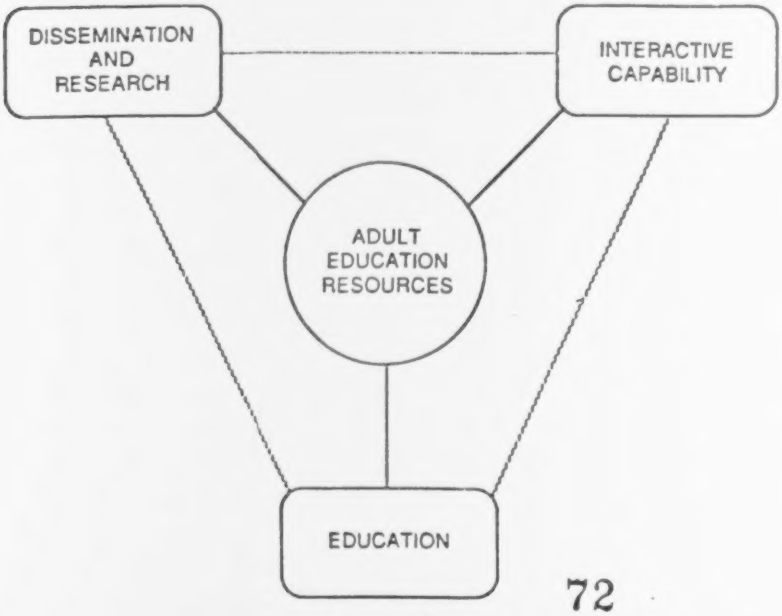
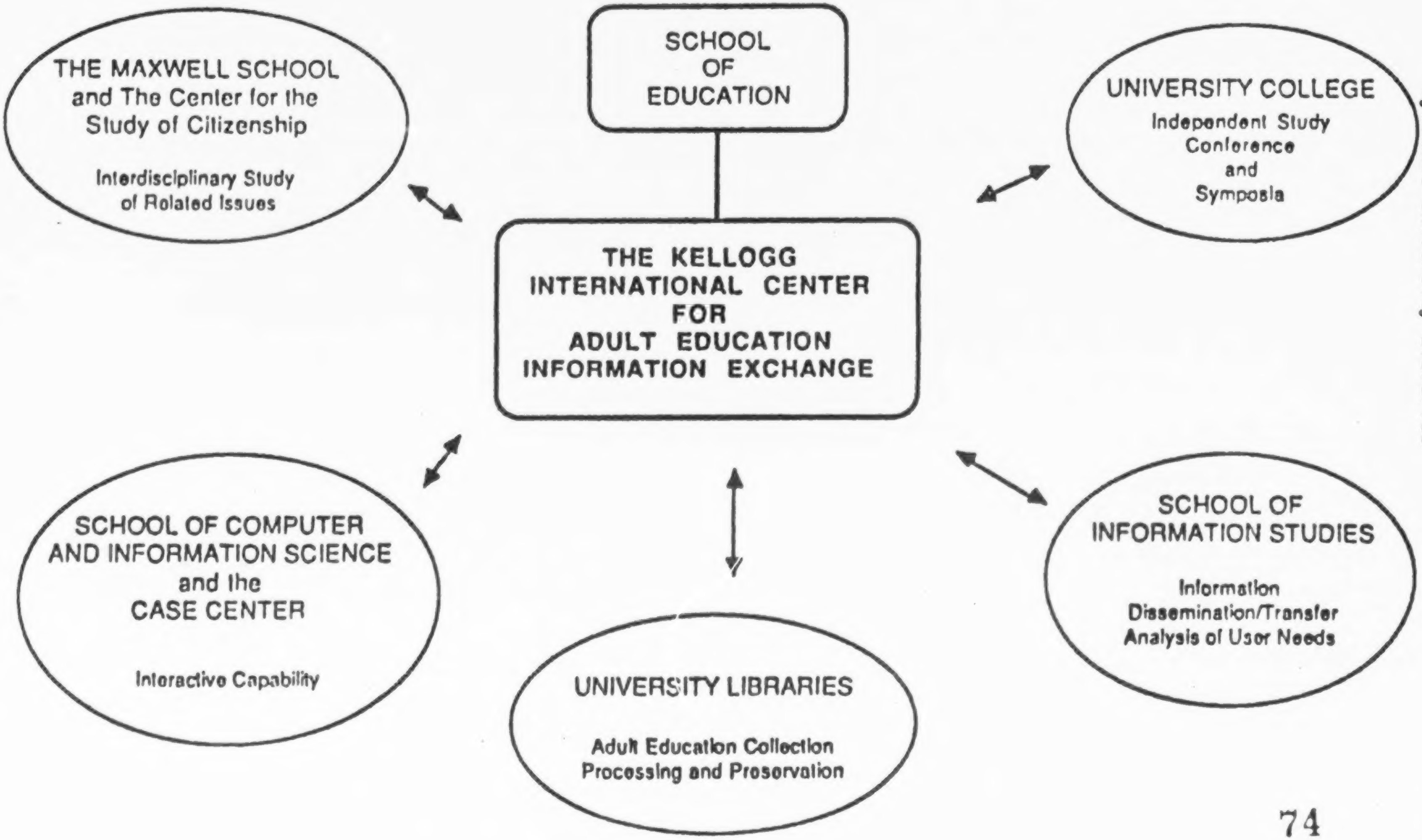


Figure 2: SUPPORTING ORGANIZATIONS


SCHOOL OF EDUCATION
THE MAXWELL SCHOOL and The Center for the Study of Citizenship
Interdisciplinary Study of Related Issues
UNIVERSITY COLLEGE
Independent Study Conference and Symposia
THE KELLOGG INTERNATIONAL CENTER FOR ADULT EDUCATION INFORMATION EXCHANGE
SCHOOL OF COMPUTER AND INFORMATION SCIENCE and the CASE CENTER
Interactive Capability
SCHOOL OF INFORMATION STUDIES
Information Dissemination/Transfer Analysis of User Needs
UNIVERSITY LIBRARIES
Adult Education Collection Processing and Preservation

Syracuse University Libraries

Project Benefits

The benefits of the proposed Project will reach far beyond the Syracuse University campus and repay manyfold Kellogg's initial investment. Below are just a few of the benefits the Kellogg Project will bring:

- Advancements in the field of adult education.
  By combining a variety of the University's talents and strengths in a number of disciplines, we will be laying the foundation for the organized base of knowledge necessary to advance the field of adult education. The University formally recognizes that the education of adults should receive as much emphasis and support as the education of youths.

- The preservation and expansion of the Collection as an outstanding resource for adult educators.
  Preserving in a digital format the Collection as a national resource for adult educators and concentrating on the acquisition of the highest-quality materials will steadily increase the Collection's level of excellence, making it even more valuable to users and providing a reputable, solid literature base for the field of adult education. In addition, every effort will be made to expand the Collection to include more international literature and professional-association materials. Syracuse will become a mecca for adult educators around the world who want to take advantage of the knowledge stored here.

- A demonstrated information-system model.
  Demonstrating a model or prototypical information system based on a comprehensive data base and knowledge base for the field of adult education will set important precedents for others. The techniques used in digitizing and disseminating information and the logic-programming tools underlying this system will provide a model for other institutions to use when faced with similar management and dissemination projects, especially in the social sciences.

- Expanded use of the Collection.
  By employing an optical-digital storage system combined with computer

technology, we will lay the foundation for making primary source materials available to adult education researchers worldwide. Users will be able to search inventories of the Collection and will eventually be able to access the full text of any document, and even obtain full printed texts without ever having to travel to Syracuse's campus.

- Enhanced communication among adult educators.
  Creating electronic networks using telecommunications technologies will facilitate interaction among national and international users, allowing them to share and discuss knowledge, ideas, feedback, experiences, and solutions to problems almost instantaneously and reassuring users that they are not operating in a vacuum, but playing an active role in shaping a young profession.

- Incorporation of the Syracuse University Libraries' adult education archival records into a national data base.
  To assure wide national visibility of and accessibility to records in our holdings, the Syracuse University Libraries will join the Archives and Manuscript Control (AMC) system of the Research Libraries Group (RLG) to catalog our Adult Education Collection into the largest and most widely used archival cataloging system in a standard archival format.

- The elimination of duplicated resources.
  Once an easily accessible center for adult education is created, other institutions and individuals will not need to purchase materials that are easily available to them through access to the Collection. This will enable them to focus their resources on their specialized collections and unique needs.

- Expertise in distance education.
  By providing an independent-study master's degree program, we can provide new opportunities to advance professional knowledge to those educators who could not otherwise pursue formal education and obtain an advanced degree because of geographical limitations.

- The exploration and sharing of new technologies.
  By taking advantage of the expertise of the various members of the advisory councils and professional staff, as well as of their experiences with and evaluations and applications of new technologies, users can be saved the valuable time that otherwise might be spent in duplication of effort. Also, as computing and telecommunications technologies become less expensive, it will be feasible to help Third World countries to set up equipment enabling them to connect with our information system.

- The illumination of important issues for adult education.
  By focusing on topics in adult education, the Center for the Study of Citizenship will create a forum for airing new ideas, diverse opinions, and various perspectives on issues that have an impact on the field of adult education.

- The ability to share information with English-speaking adult educators through AEDNET.
  By sharing through AEDNET previously unavailable information and the

latest research results in adult education with educators worldwide, knowledge can be more rapidly transferred and applied to the needs of national and international users.

PROJECT IMPLEMENTATION

This section describes the Project's operational elements: the technologies comprised by the information system, the roles of Project personnel and the advisory councils, evaluation methods, and the goals and objectives of each functional unit.

Technologies

The technology used in this Project will be aimed toward developing an optical-digital data base of adult education literature and generating a logic-programming knowledge base that will provide expanded and more flexible access to the data base than conventional information-retrieval systems can provide. The function of the system is to store and preserve the Collection and allow users to analyze critically the full text of Collection materials from nearby or remote locations, thus greatly increasing the Collection's utility to adult educators.

The system will integrate the use of optical disks with the use of logic-programming techniques. Optical-disk technology, coupled with our advanced computing powers, will make possible new ways of storing, retrieving, interacting with, and disseminating information. For the first time anywhere in the world, we will apply these methods to creating a research resource that allows users to access full-text materials conveniently through a computer system without ever leaving their desks. We will do this by:

1. Using optical-disk technology to store the complete text of archival materials as data files within the system.

2. Generating within the system a knowledge base of logic assertions about the literature that can provide flexible and convenient access to the body of the text.

3. Implementing a user/system interface that makes the system easy to use by those unfamiliar or uncomfortable with computers.

Optical disks are among the most recent developments in information storage and retrieval. In addition to having high storage capacity, they are compact, adaptable to various media, able to store high-resolution images, and resistant to wear (National Library of Canada 1985). From a budgetary viewpoint, the use of optical-digital information systems will reduce long-term costs. Although the initial cost of converting printed documents to digitized form is high, significant savings will result from the reduced physical space necessary to house these materials, and greater efficiency in retrieving information will be achieved. The National Library of Medicine and the Library of Congress are among the few libraries already using optical disks (Dominic Nghiep Cong Bui 1984).

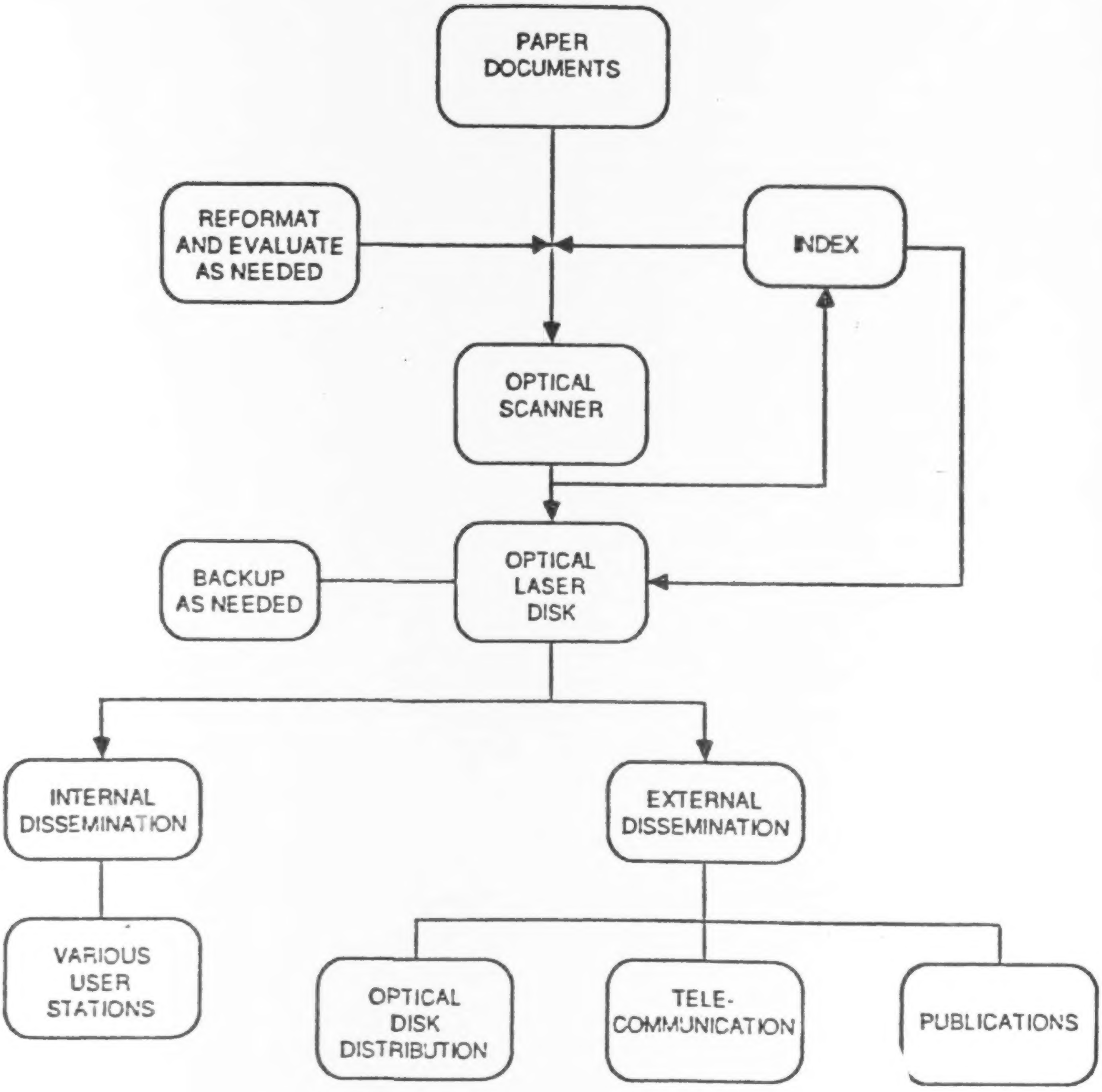
Initially, only index and catalog information will be available from a distance via the normal telephone network, which generally cannot now accommodate high-resolution images. On campus, researchers will be able to access documents contained on these disks from three multiple-information workstations, each containing an interactive control module (keyboard or similar device) with high-resolution display and print capabilities. However, we propose later to add telecommunications channels (e.g., satellite, microwave, fiber optics, coaxial) whereby users will be able to access the documents on our disks from remote places. In other words, users will be able to retrieve the full text of documents without physically coming to Syracuse's campus and having to handle them. The optical disks will enable adult educators from all over to call up the system and search the data base, review material on screen, and even obtain printed texts without leaving their desks, thus making Collection resources available to a much larger number of people in the field. It will also be possible to produce disks containing Collection documents for distribution to other information centers. (See Figure 3 for a diagram of the projected information system.)

Logic programming will play an important role in enabling adult educators to retrieve the information they seek from the Collection. We are fortunate to have on our campus some of the world's leading experts in logic programming and "fifth generation" computing, applications of which will lead to what we're calling an interactive capability for our information system. J. Alan Robinson and his associates in the School of Computer and Information Science have developed a new computer-programming system, based on logic programming, that shows promise of providing "intelligent" access to the University's Adult Education Collection. This computing environment will allow the creation of an on-line deductive knowledge base that can be queried ad hoc (i.e., in a way that closely accommodates users' information needs instead of according to preplanned formats) and that responds rapidly by means of fully automatic deductive reasoning. By such means, our large data base can be made accessible to scholars in a flexible and natural manner and with closer matches between users' information requests and system responses than conventional information-retrieval systems can provide.

The system's programs will be written in LOGLISP, a long-established and widely used language for symbolic computation. Information about the text will be entered as logical assertions, including conventional information-retrieval information. However, using the flexibility of logic assertions, we will expand the knowledge base to include additional categories of information, such as historical, political, or social descriptors of documents; identification of key sections of the text (e.g., abstracts, summaries, bibliographies, and figures and tables); and information relating text to other works (including, but not limited to, cross-references). Because of logic programming's deductive capabilities, users will then be able to search the primary data base using any combination of categories they choose.

This approach will lead to a system of great utility for users and of optimal efficiency for system-maintenance staff. First, the system will be able to provide users with more information from the knowledge base than we put into it. User queries will invoke logic computations, which will yield the implicit

Figure 3: PROJECTED INFORMATION SYSTEM


PAPER DOCUMENTS
REFORMAT AND EVALUATE AS NEEDED
INDEX
OPTICAL SCANNER
BACKUP AS NEEDED
OPTICAL LASER DISK
INTERNAL DISSEMINATION
EXTERNAL DISSEMINATION
VARIOUS USER STATIONS
OPTICAL DISK DISTRIBUTION
TELE-COMMUNICATION
PUBLICATIONS

content of the knowledge base in addition to the programmed assertions. In simpler terms, the system will be able to deduce knowledge that was not specifically programmed into it. Second, the system's flexibility will allow us to add information continually to the knowledge base. The form of new information will evolve from the users' information needs and will be completely unconstrained by the programming language.

Logic programming is a technique that is now being used in many ways by many people. However, this Project represents a completely new application that will be a model not only for adult educators, but for people in a wide variety of disciplines, particularly the social sciences. Never before has logic programming been used to design such an accessible, interactive storage and retrieval system of archival materials. The flexibility and dissemination capabilities it will provide, both for users and for archivists, will be applicable across many research and data-base management settings both in the United States and abroad.

The computer system is integral to the goals of the Project. It characterizes a technological advance that complements and enhances the goals we have established for making the Collection as widely accessible and easy to use as possible. By applying our technological expertise to the Adult Education Collection, we can establish and continually expand a truly unique body of knowledge and disseminate it to users across the nation and in other countries. In addition, we can establish a model for knowledge-base systems that will enhance the critical analysis and scholarly annotation of a literature base. Such a model will be invaluable to disciplines in the social sciences and the humanities, which are looking for computer applications that truly facilitate text-based research.

Finally, the creation of AEDNET, an electronic transmission and bulletin-board network devoted exclusively to adult educators, will provide a mechanism whereby users can communicate instantly with each other from great distances about issues vital to program development and policy. Users will also be able to share information and research findings via AEDNET, thus increasing their common base of knowledge.

Personnel

Several faculty members, specialists, and staff members will be employed to assist in the operation of the Project. (See Figures 4 and 5 for diagrams showing personnel organization; see also the budget in Appendix G for corresponding financial information.) Many of these staff positions will be needed only at the outset of the Project to process the backlog of Collection materials and put the contents onto optical disks, design and put into place the information system, set up AEDNET, and implement the interactive dissemination capabilities. Once these activities are completed, an appropriately reduced staff will be able to continue the Project's momentum. The roles of each position in the Project are described below.

Vanderbilt University Library

ENHANCING AUTOMATED INFORMATION SERVICES WITHIN THE JEAN AND ALEXANDER HEARD LIBRARY AT VANDERBILT UNIVERSITY:

A PROPOSAL TO THE PEW MEMORIAL TRUST

Vanderbilt University is committed to a library environment which fosters scholarly communication, excellence in instruction, and the expansion of resources and tools for research. In recent years the Jean and Alexander Heard Library has increasingly sought to provide automated services designed to enhance this commitment. We now propose to advance these automated services by opening up to faculty members, students, and other researchers vast amounts of information resources, both bibliographic and textual, on locally available computer facilities. This will involve mounting databases of current bibliographic information about journal articles and government documents; it will also involve increasing access to full-text resources, such as encyclopedias and dictionaries, in forms usable by many researchers simultaneously, and in compact-disc and microcomputer-based systems dispersed throughout the various units of the Library. Such new information services opportunities are becoming open to research libraries, and the Heard Library is in a unique position to take advantage of these opportunities.

CURRENT LIBRARY AUTOMATION AT VANDERBILT

Beginning this academic year (1985-86), the Jean and Alexander Heard Library has initiated an on-line integrated library system, called ACORN, designed to provide a public access catalog to most of the titles within the University's library units. The ACORN system uses Northwestern University's NOTIS software, operating on an IBM 4361 computer owned by the Heard Library and dedicated to the Library's use. The machine physically resides in the University's Computer Center and is managed by the Computer Center for the Library. The Associate Provost for Information Services and Technology and Director of the Library has organizational responsibility to the Provost for the Computer Center as well as for the Library.

The Jean and Alexander Heard Library consists of seven divisions--Central/Science, Medical, Divinity, Education, Law, Music, and Management. In addition, it also encompasses two special service units--Special Collections and University Archives, and the Television News Archive. All of these units participate in the ACORN system. From any of approximately thirty-five terminals located throughout these Library units, users can search the holdings of the entire Library. This represents a union catalog not previously available to the University. By the summer of 1986 the ACORN system should provide access to approximately ninety per cent of the Heard Library's titles. The Computer Center also plans soon to provide access to the system from locations outside the Library units through Vanderbilt's data network communications facilities.

PROJECT PERSPECTIVE

The initial implementation of the ACORN system provides bibliographic access by author, title and subject to books and to serial titles, but not bibliographic access to periodical articles and to abstracts. Neither does it provide access to the full text of important reference tools or to bibliographic information about government documents. Among the most critical next steps in the application of automated library systems at Vanderbilt will be to provide such bibliographic access to these crucial materials. At present, article-level bibliographic access is available on-line remotely (i.e. through the use of telecommunications). For example, the Heard Library currently has such telecommunications access to MEDLINE (medical articles) and to other bibliographic databases through the Dialog or BRS remote online services.

At present, however, the cost of local computer storage is falling much more rapidly than telecommunications costs are falling. This situation will increasingly allow research libraries to economize on the cost of using remote bibliographic and textual databases by mounting more data sets locally and providing access to them through campus-wide networks. When such information is locally mounted, the service will be made available at a much lower price and so should be more extensively used.

The Heard Library is making plans, therefore, to mount locally article-level bibliographic files or subsets of these files. (Subsets of files might be records for only the current year, or records focused to the specific serial titles owned by a particular Library unit, or both). The Heard Library also plans to mount locally the full text of important reference tools and to look for the availability of full-text documents where the ability to search electronically, word-by-word, will be of significant value to a substantial number of users.

D. CONGRESSIONAL INFORMATION SERVICE

1. CIS/Index is the resource and reference system for the publications of the United States Congress. Compiled since 1970, the CIS/Index contains an index and abstracts providing a number of research access points: subject, name, title, author, bill number, and publication number. The abstracts provide a brief synopsis of the publication.

2. The Index to International Statistics covers some 1,200 intergovernmental agency publications issued from some 75 sources. This Index covers periodicals, annuals, series, and special studies about population, business and financial activities, international trade, education, health, and other economic, demographic and social matters.

3. Statistical Reference Index is a guide to statistical publications published in the United States which are not, however, of government provenance. Published in an index and abstract form, the Index identifies statistics contained in publications of some 1,000 leading sources.

4. American Statistics Index allows use of the statistical output of the Federal government. Published as an index and abstract, this Index references publications of major statistical and research agencies, as well as regulatory agencies, Congressional committees, judicial offices, special councils, commissions and boards, and many other sources of economic, demographic and social statistics.

These are among the most significant databases which we anticipate being used in the Heard Library, and each of these bibliographic/reference tools points to published materials held in the Heard Library. The availability of local access to these databases would greatly enhance user accessibility to these crucial materials.

II. FULL TEXT REFERENCE TOOLS

Commonly-used reference works, dictionaries, encyclopedias, and thesauri form a second major area in the development of data resources in the Heard Library. Publishers are now beginning to make these resources available in electronic form. For example, Grolier is distributing its Academic American Encyclopedia in electronic form. This resource is a 76.5-megabyte database providing full-text searching.

Grolier is offering a site license for campus use, with home and office distribution on compact disc, a read-only media.

Our first goal is to see such electronic reference works used in about the same way as the printed versions are currently used. We plan to compare the cost of the electronic form versus that of the print form. Because we buy copies of the same encyclopedia for different locations, we might find some advantage in acquiring a single license for all the Library units if the price for that license warranted our acquiring it. We will also weigh the disadvantages occasioned by the lack of graphics against any price advantage we might incur through leasing the electronic form of the reference work.

Electronic versions of standard reference works do have some advantages over the print forms. In particular, the electronic versions are updated quarterly, and so can be made significantly more current than the print versions. The electronic version of the reference work also enables one to search for individual words and phrases, thus rendering it significantly more useful than the print version. For example, a routine search of the electronic file of the Academic American Encyclopedia (Grolier, 1985) for a topic or phrase might yield significantly more occurrences of that topic or phrase than merely looking it up in the print version. From examining the entry "gold standard" in this electronic encyclopedia, one can acquire a very general understanding of the role of the gold standard in economics and business. Persons searching through the encyclopedia's electronic file, however, would locate twenty other references to the phrase throughout the encyclopedia. These entries include discussions of persons significant in the history of the gold standard (Cleveland, McCullough, McKinley, Peel, Witte); related topics (money, exchange rate, balance of payments, inflation, free silver); and political ramifications (Democratic party, the Bland-Allison Act, the central bank). This type of free-text searching allows the location of reference works across disciplinary lines and is, thus, more helpful to users.

As more reference works become available in electronic form, the ability to access these works from one's office or dormitory room will give the electronic form a great advantage in convenience and so in usefulness. Although each desk in the University is likely to be equipped with a dictionary, few individuals have the medical, geographic, and biographic dictionaries, the encyclopedias and thesauri, the Facts on File and other frequently-used but expensive works that scholars and students find useful.

Our plans, therefore, call for us to test the usefulness of such electronic reference works and to test the cost involved in making these references available electronically.

We plan also to compare the strategy of using compact-disc-based equipment, with the accompanying probable limitations on locations, to the mounting of a widely-accessible database more centrally.

To be sure, the list of databases described above is not exhaustive. A variety of other available databases provide access to the literature pertaining to specific disciplines. As appropriate support comes available, the Heard Library will seek to acquire these specialized databases and make them available within the unit most suited to them. The Library would also provide intra-library linkages for those desiring to use those databases from elsewhere in the Library system. We also hope to encourage expanded availability of such data by convincing database producers who do not presently make their files available (e.g. H. W. Wilson and Information Access Corporation) that to do so would ultimately be cost-effective and to their own advantage.

Pilot/Demonstration Projects

THE LIBRARY OF CONGRESS

OPTICAL DISK PILOT PROGRAM

A Report on the Nonprint Project

By Carl Fleischhauer

The Experiment

The nonprint project within the Library of Congress Optical Disk Pilot Program is an experiment in this sense: if the disks help the Library perform its work better, more use of them will be made in the future. If not, little further use of them will be made. Since the disks will be of greater use for some materials and purposes than others, the project will identify the areas of greatest utility. This report will mention several facets of the experiment, but will emphasize our plans for testing the disks' contribution to research access.

Two types of disks have been created as part of the nonprint experiment: analog-signal laser videodisks for visual materials and digital audio disks for sound recordings. Laser videodisks are recorded with the NTSC (National Television Systems Committee) signal used for U.S. broadcast television. This signal offers a good quality image on a standard television screen. The compact audio disks use a digital recording system that is capable of near-perfect representation of the wave forms of sound. Digitization will permit infinite recreation of this waveform without loss or distortion.

Although it is natural and correct to perceive the experiment as an exercise in science and technology, the work extends into the realms of cataloging and intellectual property. A later section of this report discusses the interesting use of a computerized catalog to accompany disks containing still images. The matter of intellectual property has involved identifying copyright issues arising out of the storage of copyrighted materials on optical disks and their use by library patrons. The program is reproducing entire works on disk, and then publicly performs or displays them; these are rights generally accorded copyright owners.

The nonprint disks contain a large number of photographs, drawings, music, and films which are protected by copyright. The program has wrestled—and continues to wrestle—with the question of what we may reproduce, publicly perform, or display, and with the difficulties occasioned by a permission-seeking process. For older collections, the problem of identifying thousands of individual owners is massive, to say nothing of the labor of soliciting permission. In one case, we worked with the leaders of a large group of owners. As we were preparing the disks containing motion pictures and motion picture stills, the Motion Picture Association of America assisted us by arranging meetings with representatives of the major film production companies. The Optical Disk Pilot Program has received general guidance and assistance from publishers and librarians who are members of its special advisory committee.

The Disks and the Collections

The nonprint project comprises eight disks containing materials from two Library divisions. Disks one through six are laser videodisks, while disks seven and eight are digital audio disks. The list below describes the content of each disk and identifies its source, as well as noting the disk's technical features.

Laser videodisks may be manufactured in two formats: constant angular velocity (CAV) and constant linear velocity (CLV, often termed "extended play" by the entertainment industry). CAV disks are far more versatile; each frame on the disk has an individual number and may be searched or displayed at random. One may also display individual frames in slow motion. It is this capability that makes them ideal for reference use in conjunction with still image collections. Each side of a CAV disk has a motion picture or video program capacity of thirty minutes, which translates into about 54,000 individual still frames. In contrast, CLV disks do not offer random-access search, nor can they display still frames. Their advantage is greater motion picture or video program capacity. A CLV disk can contain up to sixty minutes per side.

(1) *Images from the Prints and Photographs Division I.* About 39,000 images from ten collections. The collections include the World Transportation Commission collection of hand-colored lantern slides by William H. Jackson; the Detroit Publishing Co. glass negatives, color prints, and albums; the Farm Security Administration and Office of War Information color transparencies; the Abdul-Hamid II Collection; the Yanker Poster Collection; and a group of selected architectural drawings.

None of these images is protected by copyright in the United States. This is a single-sided, still-image disk in the CAV format. It may be played in two modes. First, there is an automated, menu-driven program for the Sony disk player; this program offers general, collections-level access to the content of the disk. Such an automated program—resident on the disk and loaded and run by the disk player—is known in the industry as a "level two program." Although not fully operational at this writing, the disk may also be played in conjunction with a separate microcomputer, known in the industry as a "level three program." A subsequent section of this paper elaborates on the level three display system.

(2) *Images from the Prints and Photographs Division II.* About 10,000 images from three collections. The collections are the Cabinet of American Illustration, the American Cartoon Drawings Collection, and the Architectural Drawings Collection.

Many of the images are protected by copyright; thus, what we may do with these images is limited. One copy of the disk is being displayed in the Prints and Photographs Reading Room. This is a single-sided, still-image disk in the CAV format, with a level two program for the Sony disk player. Data about this disk will be available in the division's microcomputer database, and this disk will be displayed in the level three system.

(3) *Motion Picture Stills*. From the Motion Picture, Broadcasting and Recorded Sound Division. About 90,000 motion picture publicity stills.

Many of these images are protected by copyright; special arrangements have been made to permit their reproduction and public display for the duration of the pilot program. This is a two-sided, still-image disk in the CAV format. Each side contains a table of contents arranged alphabetically by film title. The table provides the frame number addresses for groups of stills, which may then be searched by use of the player's remote control device. This mode of operation is known in the industry as a "level one program."

(4) *Color Test Disk*. From the Motion Picture, Broadcasting and Recorded Sound Division. A selection of seven films and film segments with special color preservation problems. Titles include *Women of Tomorrow*, *Unusual Occupations*, *With the Marines at Tarawa*, *Alaskan Grandeur*, *Jewel of the Baltic*, *A Hatful of Dreams*, and *Taxi Driver*.

All of these films except *With the Marines at Tarawa* are protected by copyright. Special arrangements have been made to permit their reproduction and public performance for the duration of the pilot program; in order to simplify the matter of identifying copyright owners and seeking permission, the soundtracks have been deleted from all titles except *With the Marines at Tarawa*. This two-sided disk is in the CLV format.

(5) *Early Films from the Paper Print Collection*. From the Motion Picture, Broadcasting and Recorded Sound Division. A selection of about 30 titles from the 3,000-item paper print collection. These films focus on President William McKinley, his assassination, and the 1901 Pan-American Exposition.

None of these films is protected by copyright, and they may be reproduced and displayed to the public. The disk is to be produced during 1985, and will be a two-sided CAV disk for non-automated, level one play.

(6) *Teleproductions*. From the Motion Picture, Broadcasting and Recorded Sound Division. Two CBS television network newscasts from the 1976 Fourth of July weekend.

The programs are protected by copyright; special arrangements permit their reproduction and display to the public. This two-sided disk is in the CLV format.

(7) and (8) *Summer Chamber Festival 1982* (Works composed by Oliver Messiaen, performed by Alexis Galperine, Loren Kitt, Carter Brey, and Hugh Wolff) and *Summer Chamber Festival 1983* (Works composed by Béla Bartók, performed by Richard Young and Junko Ohtsu, and Robert Schumann, performed by Hugh Wolff, Alexis Galperine, Junko Ohtsu, Miles Hoffman, and Carter Brey). From the Motion Picture, Broadcasting and Recorded Sound Division. Two concerts recorded in the Library's Coolidge Auditorium.

Two of the musical compositions are protected by copyright; special arrangements permit their reproduction on this disk. The disks are available for listening. The disks are compact digital audio disks ("CDs"). The disk player offers access to individual movements within the selections.

Preservation

If the disks live up to their promise, they will help preserve the Library's collections in four ways. First, the audio and teleproductions disks (6, 7, and 8) reproduce sound and image with sufficient quality to permit the Library to consider discarding the originals. Second, the still-image disks (1, 2, and 3) will substantially decrease the wear and tear on collections caused by handling. Third, the disk containing contemporary motion pictures (4) offers a kind of preservation compromise: although it does not reproduce the film perfectly, it offers the Library a very interesting preservation option at this time. And fourth, the production of the disk containing early films from the paper print collection (5) required the Library to restore and enhance the films in order to put them on disk.

The two audio disks (7 and 8) and the *Teleproductions* videodisk (6) reproduce the material they represent without significant loss of quality. This results from the near-perfect replication offered by the digital audio signal, and the fact that the teleproductions are simply transferred from one video format (¾-inch U-matic network logging copies) to another (the 1-inch video premaster and the videodisk itself). For the teleproductions, there will be a slight loss of quality incurred by the transfer ("generation loss"), but this loss would be incurred regardless of the choice of preservation medium. If the videodisks have a long life, this transfer will be the final one.

The NTSC television signal, however, cannot faithfully reproduce the fine detail of an architectural drawing or a large-format photograph. The disks with drawings and photographs (1, 2, and 3) will serve preservation by reducing the need to handle the original materials. Many of the original items are works of art and historic documents with individual qualities and intrinsic value. No form of reproduction—no matter how excellent—would lead to their being discarded.

The still images for these disks (1, 2, and 3) have been photographed on motion picture film, which is transferred to video when the disk is mastered. For this edition of disks, the original items could have been photographed on 16mm film, which provides resolution sufficient to create a high quality NTSC signal. Instead, a more expensive 35mm film was used. Its higher resolution means that the motion picture film itself offers a reasonable preservation copy of all of the images. This high quality film copy, of course, may be retransferred to future high definition video formats without the need to rephotograph the collections.

Vidodisks may be able to retain image color more accurately than some motion picture film stocks. Not only do many film prints suffer from color shifts, but their negatives may also change color in time—lack of color stability is a severe problem for a film archive. Thus, even though the NTSC television signal is incapable of reproducing every nuance of a motion picture image, the *Color Test Disk* (4) will test the applicability of the videodisk form for motion picture color preservation. For certain films, such disks may be the only feasible preservation medium available.

The *Color Test Disk* will serve an additional preservation-related function. The films it contains were printed in a variety of color processes: Technicolor, Magnacolor, Eastman Color, Anscochrome, and Cinecolor. The film prints will fade at different rates. By using the disk as a yardstick, the Library may be able to identify the processes which fade most quickly and establish preservation priorities.

The second motion picture videodisk (5) represents another aspect of preservation, perhaps better called restoration. The paper print collection dates from the very earliest days of film and film copyright. The Library's original copies of this historically important body of work are literally on paper rather than on film. The imagery is unstable in various ways, and the speed (frame rate) fluctuates. The transfer to videodisk involves sophisticated analysis of the footage, and frame-by-frame image correction.

Research Access

The disks' promise for improved research access varies from case to case. Interestingly, the pilot program disks which offer the least contribution to direct preservation will make the greatest contribution to research, and vice versa. The three still-image disks contain large numbers of items representing complete collections. They will serve researchers well, and will provide the best opportunities for testing. In contrast, the motion picture, teleproduction, and audio disks offer only a handful of items, and they will not serve researchers in an effective way. For example, there are only two television broadcasts on disk. This would be like having only two issues of the *New York Times* on microfilm. Such a microfilm could not be tested for research value in a reading room. Only mere chance would bring a reader into the room, seeking those particular items.

Disks with Still Images. For these disks, two significantly different retrieval methods are being tested. The Prints and Photographs Division system employs a microcomputer which will be linked to the disk player. A full description of the system is provided in a longer paper by Elisabeth Betz Parker to be published in the December 1985 issue of *Information Technology and Libraries.* The Motion Picture, Broadcasting and Recorded Sound Division stills disk offers access by means of a simple, alphabetical title index.

The Prints and Photographs Division microcomputer contains data records ("captions") offered to researchers by means of the BRS/Search (Bibliographic Retrieval System) software, a readily available commercial package. The captions—which may be thought of as abbreviated MARC (Machine-Readable Cataloging) records—list information about a picture, including the name of the creator, the title, subject terms, date of creation, copyright information, storage location of the original image, existence of copy negatives, and any additional information that may be available. Of course, they also identify the videodisk frame number where the picture may be found. They are *item-level* records; they coexist with *collection-level* full MARC records. This two-tiered database has provided an interesting model for consideration by other Library special collections. At this writing, about one-quarter of the captions has been prepared.

If a researcher wishes to locate the work of a particular photographer, a scene including a particular subject, the date of depiction, or any other aspect of an image likely to be mentioned in a caption, the system will search the database. This is like conducting a search for a book on the Library's computerized catalog. In the interim display system, the frame number must be keyed in by the user on the player's remote control, thus causing the image to be displayed. In the final, level three display system, the researcher will press a function key and the computer will automatically instruct the player to display the image.

The system also permits a researcher to look at pictures first, and seek captions later. One may browse through the images on the disk and, if interesting examples are found, instruct the computer to find the caption. In the interim mode, this search is accomplished by searching for the disk frame number, which is a part of the data record. In the final display system, the search will be performed automatically by pressing a function key.

The final setup for this system will test the overlay, or superimposition, of captions on still images. This is in contrast to a two-screen display system, with data on one screen and the image on the other. The advantage of the overlay method is that both the visual image and the text are displayed on a single screen. This not only integrates the data, but also reduces the amount of hardware required for a user station. The overlayed captions, of course, would only be displayed at the researcher's request. Two methods for overlaying will be tested: a method that uses the NTSC signal and an RGB ("red, green, blue") method which beams the text information directly onto the picture tube for greater clarity.

The disadvantage of the caption overlay is the risk that the combined text and image may be cluttered and the text hard to read. In addition, the low resolution of the television picture tube precludes the display of the entire data record. The testing will alternate between an equipment configuration which offers the overlay display and one which offers the two-screen display. An advantage of the two-screen method is that the text can be displayed on a high-resolution computer-type screen, which offers a larger number of readable characters, and which will not clutter the visual image. Both images can be consulted simultaneously.

The interim Prints and Photographs Division display system includes only an image printer; the final system will feature both an image and data printer.

The *Motion Picture Stills* disk contains groups of publicity photographs for about 6,700 motion pictures. A table of contents on the disk and in an accompanying booklet lists the titles in alphabetical order and provides their frame addresses; there is no computerized database. The booklet also contains all of the copyright notices from the original stills. The Library's motion picture collection is indexed by title (there is no full catalog with subject-term access) and this disk will offer researchers the stills in the same way. Access to the stills on the disk, of course, is greatly improved because search time is reduced compared to the time spent finding and pulling items stored in vertical files.

With this arrangement the hardware requirements are slight: only a player and monitor are needed. And this level one disk can be used with far less help than the more complex Prints and Photographs Division level three system, which depends upon some on-site reference staff assistance.

Disks with Motion Picture and Teleproductions. Videodisks containing motion pictures—whether produced by this project, by the motion picture industry for home entertainment, or by the Library in the future—will not have the substantial impact on access and use that the still-image disks have. Most videodisks of films contain only one or two titles. With disks containing tens of thousands of still images, finding one's place is a critical problem. With a disk containing a few film titles, the problem is one of finding the disk, not one's place on the disk. The retrieval issue is thus a cataloging issue. The disk may take the place of (or supplement) the cans of film, but both items are sought and retrieved from a general catalog of motion pictures in the Library's collection.

For the Library, the question with respect to research use becomes one of acceptability. To what degree will the image quality of a videodisk inhibit use? To what degree are features such as still-framing and frame searching desirable? In contrast to a fragile film print, a disk may be viewed repeatedly. Do researchers feel that the added value of repeated viewing makes up for lower image quality? The severely limited research scope of the two motion picture disks being created by the optical disk project prevents them from offering a solid basis for asking and answering these questions. But the topics are of interest and, although firm plans have not been made, there has been discussion of soliciting preliminary reactions to these questions from selected users.

The disk that contains selections from the turn-of-the-century paper print collection will explore the role videodisks can play with large bodies of archival material. Here a videodisk would not only preserve footage, but could also make material more accessible for cataloging as well as use. The early films disk will also experiment with the display of film exposed at frame rates which do not conform to today's norms. The frame rates for the paper prints vary, but tend to average from 12 to 18 frames per second, about half the rate of contemporary motion pictures and video. On one side of the disk each disk frame will represent two film frames. When played at the usual 30 frame-per-second playback speed for videodisks, action on the screen should appear more or less normal. On the other side of the disk, each film frame will occupy one disk frame. In order to represent action properly, a viewer will have to view this side of the disk with the player set for slow motion. This may cause a bit of flickering, but will allow twice as much footage to fit on the disk.

The comments offered in regard to videodisks of motion pictures generally apply to the disk with television broadcasts. Retrieval and access are accomplished by a general catalog and do not require within-disk searches.

The Digital Audio Disks. Digital audio disks can contain up to 70 minutes of program material; neither of the pilot program's two audio disks will contain a large number of items. For reasons similar to those noted in the discussion of videodisks of motion pictures and teleproductions, the problem of access is a problem of cataloging. A good general catalog of the Library's recorded sound collections will enable a researcher to find an item.

The disk players, however, lend themselves to sophisticated automated playback. A music researcher, for example, might instruct the player to repeat a musical phrase again and again and thereby facilitate comprehension and transcription. The experiment may examine the potential of this capability in a limited test.

The low audio storage capacity of compact disks is a drawback for an institution with vast collections of sound recordings, and the Library will continue to experiment with other ways to preserve sound recordings on optical disks. Discussions are under way concerning testing a specially formatted videodisk that offers high audio capacity, albeit with a limited frequency range. Such a format may be suitable for some of the Library's historical collections.

Evaluation Testing

Program Evaluation. The experiment will be evaluated by a combination of informal and formal techniques. The most important informal evaluation will come from the reference staff. These librarians work with the patrons who use the disks, and they are perfectly situated to observe how well the system performs. In addition, they can compare this performance with their prior experience with the same collections.

The formal evaluation will focus on the richest testing opportunity: the system in the Prints and Photographs Division Reading Room. The evaluation will employ questionnaires and interviews with selected patrons and reference staff. A slight amount of data will be derived from user tallies collected by the display system's computer.

Technical Evaluation of the Disks. The Library's Preservation Office will test the project's disks to try to determine the likely life of the disks and identify preferred storage environment and media.

It is worth noting that analog-recorded information suffers generation loss each time it is copied. This is in contrast to digital information, such as that on the sound recordings listed above and on the text-image recordings produced by the print project of the Optical Disk Pilot Program. The implication of generation loss is that the *mastering materials*, particularly the videotape "premaster" and the disk master itself (on a glass base), must be protected in order to permit recreation of a disk. The Preservation Office plans to test mastering materials as a key part of its effort.

Dissemination

The Library of Congress hopes that the Optical Disk Program will serve as a model for other institutions who seek practical solutions to problems of access, security, collections control, and preservation. The program will disseminate information by staging presentations and demonstrations for Library staff and interested outsiders, and by preparing publications and news releases to communicate with a broad audience. At this writing, about 1,000 visitors have seen demonstrations of the Prints and Photographs Division system. Within the limits of copyright arrangements, and after completion of the pilot program, the Library hopes to find ways to permit other libraries and research institutions to share some of the disks and data.

Library of Congress

LIBRARY OF CONGRESS OPTICAL DISK PILOT PROGRAM

COMPARATIVE CHARACTERISTICS OF ANALOG VIDEODISCS AND DIGITAL DISKS AS USED AT THE LIBRARY OF CONGRESS

ANALOG VIDEODISC	DIGITAL DISK
Technology	
Commercially available home entertainment type discs.	Computer mass storage and retrieval; digital disk systems entirely computer driven.
Uses off the shelf technology.	Technology is evolving.
Low resolution display terminals/monitors linked to NSTC television technology of 525 lines per screen.	High resolution technology. For example, LC system will have a 300x300 lines per inch, scanning, storage, and printing capability.
Storage capacity	
54,000 TV resolution images on one side of a 12 inch disk.	10,000-15,000 high resolution pages on one side of a 12 inch disk depending on the amount of error correction coding used.
Physical disk	
Round, various diameters, shows color diffraction patterns.	Round, various diameters, shows color diffraction patterns.
Surface pits of an indefinitely variable length corresponding to frequency modulation and coding of video information.	Surface of uniformly spaced bumps or bump free areas which correspond to fixed length bit information; pre-grooved to assist in tracking.
Disk Production	
Produced at special mastering facilities in the US or Japan by mastering one disk using a laser and replicating this disk by compression or injection molding; all copies pressed from the same master.	Produced by direct read after write or DRAW type technology, i. e., disk is laser written and verified on site, in 'real time'.
Access/Retrieval	
Random access.	Random access.
First level access/retrieval provided by player and disk; higher level retrieval provided by associated programmed microcomputers.	All operations computer driven: input, recording, retrieval; allows for complex retrieval when programmed.

ANALOG VIDEODISC	DIGITAL DISK
Technical features	
Limited image processing and enhancement capabilities.	Excellent image processing and enhancement features.
Color routinely available.	Color costly since each color must be separated at scanning and then stored and retrieved separately.
Analog coding not compatible with digital coding; play units for either system not interchangeable.	Digital coding is not compatible with analog coding; play units for either system not interchangeable.
Halftone images easily represented with fidelity to the original.	Greater storage required for halftone images because of the large amount of space used in representing the gray scale in the halftone image. However, through using an 'ordered dithering' process, halftone can be approximated with fewer bits.
Material stored	
Most useful for images where color or motion sequences are important.	Used where a high resolution display is needed.
Preservation characteristics	
Unknown life expectancy.	Better known life expectancy due to dynamic error correction and detection and programmed disk rewrite.
No error correction capability.	Both error correction and detection built into the system; data quality electronically monitored.
Transfer frequently causes data loss and hence poorer quality images on new copy.	Transfer causes no data loss.

THE LIBRARY OF CONGRESS

OPTICAL DISK PILOT PROGRAM

The Library of Congress holds the largest body of recorded knowledge in the world. Its 80 million item collection grows at a rate of 7,000 new items per day, but a significant number of older items are in an advanced state of deterioration. As a result, the Library is pursuing major research efforts to discover new ways to store, preserve and access its vast resources.

In late 1982, the Library began research in the use of optical disk technology for image preservation and retrieval. During the Optical Disk Pilot Program, the Library is studying the use of this technology for the preservation and management of its collections and evaluating the cost and benefits of such technology when used in a production setting.

Because optical disk technology affords extremely high-density storage of information, it is a potentially efficient and economical way to store and retrieve images. A one-sided 12-inch digital disk can store between 3,000 and 10,000 pages of text depending on the distribution of black and white on the pages. One side of a 12-inch analog video disk can store up to 54,000 pictures. Optical storage also avoids wear resulting from continuous disk usage since only a beam of light touches the medium during playback. In addition, the lifetime of the information on a digital optical disk can be extended indefinitely if the original disk is copied before it begins to deteriorate seriously.

The present Pilot Program was stimulated in part by the Library's early exploration of optical disk technology for use in the Cataloging Distribution Service, where there was an early attempt to scan catalog cards for digital storage and for the subsequent reprinting of copies of individual cards on demand. Such experimentation and suggestions from researchers at the forefront of the technology around the nation prompted The Deputy Librarian of Congress to appoint an Optical Disk Storage Technology Committee. This group used information obtained from manufacturers to define equipment and disk specifications, and solicited bids. Two contracts resulted: one with Integrated Automation to supply an experimental optical digital disk system and the other with SONY Video Communication Products Co. to produce six video disks and two compact audio disks, and to supply several commercial players and monitors.

The Optical Disk Pilot Program, thus, has two aspects. Print or textual materials are stored on digital optical disks. Non-print or pictorial materials are stored on analog optical disks, or, as they are commonly known, video disks.

PRINT MATERIALS

A representative group of printed library materials was selected to be scanned and stored on optical disks. These included periodicals, government publications, manuscripts, maps and sheet music. The digital disk project gave initial emphasis to high-use current materials from the Congressional Research Service's Selective Dissemination of Information (SDI) service (consisting of periodical articles and government documents on public affairs topics). In mid-1986, the *Congressional Record* for the 99th Congress and a small manuscript collection were also scanned and made available for public use. Additional collections intended for inclusion in the pilot program are entire issues of serials with article and issue level access, as well as collections of maps and sheet music.

Processing this material involves physical preparation, the creation of an index record, scanning, quality review, and writing to optical disks. In the document preparation unit, index information for each document to be scanned is entered via computer terminals into the Library of Congress Information System (LOCIS) database. A document number is assigned that later will be associated with information on the location of the document on an optical disk. Documents are then scanned and digitized at 300 lines per inch and temporarily stored on a magnetic disk. After a full quality review, the information is transferred by a laser writing process to an optical disk.

To retrieve the information, the user signs on to one of the available files at a high resolution terminal, performs a search, and identifies an appropriate citation. Upon receiving the command to display the text, the Library's computer links the request to data identifying the location of the image encoded on the disk, selects the appropriate optical disk from the jukebox carousel, downloads the image data into temporary magnetic disk storage, and displays on the terminal screen the image of the first page of the document. It is then possible to page forward and backward through the document. The user can print the images page by page at convenience printers associated with the user stations or request off-line prints of the complete document for delivery within 24 hours.

To obtain detailed information revealing user concerns, the system provides users with an online questionnaire. The Library is able to monitor actual patterns of use by making a confidential computer record of user transactions.

Finally, the Library is putting disks to laboratory tests in an attempt to determine their likely useful life and participating in national and international efforts to set standards for optical disks of all kinds.

NON-PRINT MATERIALS

A variety of visual media have been recorded on analog laser video disks to test the utility of this technology for preservation of pictorial and sound materials and improved access to such collections by researchers. Two video disks contain still images of the following items from the Prints and Photographs Division: 875 lantern slides of views from around the world photographed on the World Transportation Commission Trip (1894–1896); 25,000 photonegatives (displayed as positives) from the Detroit Publishing Co. Glass Photonegative Collection; 2,900 transparencies from the Detroit Publishing's Glass Transparency Collection; 58 colored large format photoprints from the Detroit Publishing Co.; 241 lithographically colored photoprints from Detroit Publishing albums; the Abdul-Hamid Collection of 1,825 black-and-white photos of the Ottoman Empire; 645 color slides and transparencies from the U.S. Farm Security Administration; 965 color slides and transparencies from the U.S. Office of War Information; 3,615 political and propaganda posters from the Yanker Collection; and 440 items (plans, details, sections, renderings, measured drawings, etc.) from the Master Architectural Drawings Collection for the Library's Jefferson and Adams Buildings and from the Historic American Buildings Survey. A third still image video disk contains 90,000 motion picture publicity stills from the Motion Picture, Broadcasting and Recorded Sound Division. The three moving image video disks contain films or segments of films with special color preservation problems, films from the earliest days of the medium, and two television news broadcasts. Two compact digital audio disks offer concerts recorded in the Library's Coolidge Auditorium.

All of the still pictures were filmed on 35mm motion picture color film and transferred to videotape, which was used to produce a master video disk. User stations consist of video disk players and standard color television monitors where the images are displayed in color or black-and-white. A caption database has been created for the items from the Prints and Photographs collections enabling the user to perform free-text and controlled vocabulary searches for images stored on disk. Conversely, a patron browsing images can call up the corresponding caption information from the database. Video disk stations in the Prints and Photographs Division are also equipped with thermal video image printers which provide for ready reference black-and-white scratch copies of the images displayed.

The Library is actively studying the results gained from the pilot program so far. It is looking at the impact of the technology on organization, collection development, collection service, the distribution of information, preservation and budget. It intends to phase in the new technology in all appropriate activities throughout the organization.

For more information, write or call the Library of Congress Office of Planning and Development, Washington, D.C. 20540 (202-287-6233).

September 1986

Massachusetts Institute of Technology

One-Page Description of Barker Engineering Library Demonstration Site

M.I.T. Libraries

The Barker Engineering Library has been designated as the demonstration site for Digital Equipment Corporation's new CD-ROM bibliographic searching system. The system includes various databases on 4.7" compact disks; two databases are currently available: the Computers, Communications, and Electronics section of the National Technical Information Service and the Health and Safety in Chemistry section of Chemical Abstracts.

CD-ROMs (compact optical discs - read only memory) are being used by database publishers as a high-density medium for the distribution of databases. Libraries can subscribe to these CD-ROM databases much as they do to printed materials. CD-ROM databases are expected to have a significant impnact on the database world in 1986. Unlike remote databases that charge for connect time, locally held databases can be searched as much as desired with no additional charges.

Users can sit down under the Barker dome and search for articles and reports within the scope of the available databases. Access points to the databases include author, title, subject, and keyword.

The system will be available in Barker for public use through June. Users can ask for directions and assistance at the Barker Library Information Desk. Comments and suggestions on the system are welcome; users will be asked to fill out a short questionnaire.

M.I.T. Libraries

SET-UP PROCEDURES FOR THE CD-ROM SYSTEM

1) Get the CD-ROM player, keyboard, black power cords, and green power pack from the filing cabinet at the Information Desk. Get the NTIS disk from Circulation.

2) Place the CD-ROM player on the top shelf of the desk; attach the white cord and tighten the two screws. Attach the cord from keyboard to the back of the CRT.

3) Plug the green power pack into the outlet under the desk. Connect the three black power cords by plugging one end of each into the green power pack (note: may already be done) and by plugging the other ends into the backs of the CD-ROM player, CRT, and printer. Plug the black power cord coming from the back of the Micro-Vax unit into the green power pack.

4) Turn on the power in this **EXACT** order:

- **a)** Turn the power on by hitting the left/right toggle switch on the green power pack. Two lights will come on.
- **b)** Turn on the CD-ROM player by pushing in the round, black button at the back of the machine. A green light will come on.
- **c)** Turn on the CRT by hitting the left/right toggle at the back of the machine. A green light will come on.
- **d)** Turn on the printer by hitting the up/down switch at the back of the machine. The red light will come on and must be under "power" and not under "fault".
- **e)** Turn on the Micro-Vax unit by hitting the up/down toggle switch to the right of the words "Micro-Vax I". There are four buttons below this toggle switch; these buttons push in and out and they should all be in the out position (note: normally this is not a problem unless someone has been fooling with the machine).

5) Put the NTIS disk in the CD-ROM player. The door opens and closes easily by touch.

6) Enter the date and time as requested. For example, **23 feb 1986 12:15.** From this point on, please **BE PATIENT**, this is not a fast system.

7) After 2-3 minutes you receive a message that begins with a blinking cursor and the word **SYSTEM** followed by a line or two, hit **RETURN.**

8) When requested to enter your username, enter the word **USER.** When requested to enter your password, enter the word **PRIMAVERA.**

9) At the **$** enter **@NTISEL1.COM** and hit **RETURN.** Please **BE PATIENT**, this is not a fast process. Ignore all messages until you receive:

ENTER YOUR REQUEST
1/

10) Type **QUIT** after the **1/** and wait until the **$** appears. Remove the disk from the player and return it to Circulation. The system is now booted and ready for its first user of the day.

(Version Dated 2-19-86 -- RH)

M.I.T. Libraries

CD-ROM SEARCHING IN BARKER LIBRARY

As you may already know, Barker Libray has been designated as the demonstration site for DEC's new CD-ROM bibliographic searching system. The system includes a Micro-Vax I, compact disk reader, and two 4.7" compact disks. Each disk contains a sample database: One is the Computers, Communications, and Electronics portion of the NTIS database, and the other is the Health and Safety in Chemistry portion of the Chemical Abstracts database.

Carole Schildhauer and Rich Hines of the Barker staff will be giving four demonstrations of the system over the next few weeks. Each demonstration will last about an hour, and will consist of a short lecture and a number of sample searches. We would appreciate it if attendees would take some time working with the system, either immediately following the demonstration, or at a later date, and then fill out a questionnaire concerning their reactions to the system.

A list of dates and times for the demonstrations are listed below. If you would like to attend, please call Deborah Smith at x3-7742 to reserve a place.

March 24, 1986	(Monday)	3-4 P.M.
March 26, 1986	(Wednesday)	10-11 A.M.
April 1, 1986	(Tuesday)	1-2 P.M.
April 4, 1986	(Friday)	11-12 A.M.

Massachusetts Institute of Technology

BARKER ENGINEERING LIBRARY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ROOM 10-500
CAMBRIDGE, MA 02139

INSTRUCTIONS FOR THE USE OF THE CD-ROM SEARCHING SYSTEM

1) The Barker Engineering Library, in cooperation with the Digital Equipment Corporation, is making available selected portions of the National Technical Information Service and Chemical Abstracts databases. The NTIS database (52,212 items) covers computers, communications & electronics; the CA database (86919 items) covers health & safety in chemistry.

2) This system will allow you to do your own author, title, and subject searches within the limited areas provided by the databases. The advantage of this system is that the information is on disk and therefore online cost is not an issue.

3) Using the system is very easy once you understand the command language. Each disk comes with an explanatory booklet that explains the commands, syntax, fields, indexing structure, etc. You should read this booklet before searching. **We strongly suggest that you ask at the Information Desk for a quick orientation to the system.**

4) Logging on is simple! Put the disk in the player; it opens and closes easily by touch. Make sure that you have a **$** prompt, then look at the CD-ROM player; if there is a red light on, type **dismount dub0.** At the next **$** prompt type **@NTISEL1** for the NTIS database or **@CASAFE1** for the CA database. If there is no no red light on, just type **@NTISEL1** or **@CASAFE1**. Please be patient, booting up will take a few moments. You are ready to begin your search when you receive the message:

```
ENTER YOUR REQUEST
   1/
```

5) Logging off is also simple! When done type **Quit.** When you receive the **$** prompt type **dismount dub0**, then remove the disk and return it to circulation.

6) Hints: While using the system please be **patient.** It often takes quite a few seconds to complete a search. Also note that there is a **Hold-Screen** button in the upper left hand side of the keyboard; use this to stop scrolling of the citations while they are being displayed.

7) PLEASE SEE NEXT PAGE FOR SAMPLE SEARCHES.

8) Sample Searches (Your input is in bold, results are in non-bold):

a) Looking for information about computers and biomedicine:

1/ Find computers and satellites

```
1107 1/ computers
  34 2/ satellites
   3 3/ computers and satellites
```

3/ Display=3 All for 1-3

Note: At this point all three citations from set #3 would be displayed; you will be prompted between citations to enter any necessary commands to move from one citation to another. Use the Hold-Screen button to stop scrolling of citations that are longer than one screen.

b) Looking for information about an author, Marvin Minsky:

4/ Look au=Minsky, M

```
.ITEMS          TERMS

A     2         AU= Minor, E.E.
B     1         AU= Minor, Lewis
C     8         AU= Minott, P.O.
***** YOUR TERM *****
D     5         AU= Minsky, M.L.
E     1         AU= Minsky, Marvin
F     1         AU= Minsky, Marvin L.
```

4/ d,e,f

5/ Display=4 TAN for 1-7

Note: You have created a set of 7 items (d to f); now you are displaying all 7 citations in a short format (title and accession #). See the person at the Information Desk for a complete explanation of the commands and fields involved in these searches.

Massachusetts Institute of Technology

MicroBASIS COMMAND SUMMARY

COMMAND	DIALOG EQUIVALENT	SYNTAX	EXAMPLE
LOOK	Expand	Look <Prefix>=<Value>	Look ti=electric
		Look <Prefix>=<Stem>*	Look au=smith*
FIND	SS	Find <Prefix>=<Value>	Find ab=nuclear
		Find <Prefix1>;<Prefix2> [;...<Prefix n>]= (<Value>)	Find ab;ti=(solar)
		Find <Doc Set #>	Find 4
		Find <Any of Above> <Boolean Operator> <Any of Above>	Find 4 and ti=energy
DISPLAY	Type	Display[=Doc Set #] <Field Name(s)> FOR <Item #s>	Display=1 all FOR 5 Display ti, ab FOR 2-4
PRINT	Print	Print[=Doc Set #] <Field Name(s)> FOR <Item #s>	Print=3 all FOR 1,3-5
LIST	Display Sets	List [Argument]	List List Fields
MORE	Page	More	More
?	?	?<Topic>	?Commands
*	?	[Argument]=<Stem>*	Find ti=comput*
QUIT	Logoff	Quit	Quit

IMPORTANT NOTE: Consult MicroBASIS: a Primer in RH's Office for a complete list of commands and examples.

(Version Dated 2-19-86 -- RH)

Massachusetts Institute of Technology

PLEASE HELP THE MIT LIBRARIES AND THE DIGITIAL EQUIPMENT CORPORATION EVALUATE OUR NEW CD-ROM BIBLIOGRAPHIC SEARCHING SYSTEM BY COMPLETING THIS QUESTIONNAIRE

1. Are you: ☐ MIT ☐ Non-MIT
   ☐ Undergraduate ☐ Graduate
   Department: ______________________________
   ☐ Faculty ☐ Research Staff
   Department/Lab ______________________________
   ☐ Library Staff
   Library: ______________________________
   ☐ Other
   Affiliation/Department:
   ____

2. Which database did you use? ______________________________

3. How long did you sit in front of the terminal before you began your search of the database: ______________________________

4. How long did you search for: ______________________________

5. Did you use any documentation to assist you: ______________

6. If so, which of the following:

   ☐ Two-page introductory instructions
   ☐ Mini user's guide in diskette box
   ☐ Quick reference sign on desk shelf

   Your Comments about the documentation: ______________________
   ____
   ____

7. Did you require assistance in using the system: __________

8. Did you have problems with the layout of information on the screen:
   ____

9. Did you have any problems understanding the commands necessary to use the system:
   ____

10. Did you use the Help function: ______________________________

11. If so, your comments: ______________________________
    ____

12. Were you satisfied with the speed of information retrieval and display: ______________________________

(PLEASE SEE OVER)

13.Was the database complete enough for your purposes: ______

14.If not, what was missing: ______________________________

____

15.What type of searching were you doing:

☐ Author
☐ Title
☐ Subject
☐ Other

16.If you had the chance, how often would you use the system:

____

17.What other databases, or types of information would you like to see on the system:

____

____

18.Have you ever done online searching before: __________

19.If so, on what systems: ______________________________

____

20.Use this space to make any other comments you may wish to:

____

____

____

____

____

____

THANK YOU! PLEASE LEAVE THIS FORM AlONG WITH THE COMPACT DISK AT THE CIRCULATION DESK.

CD-ROM
BARKER LIBRARY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ROOM 10-500
CAMBRIDGE, MA 02139

Evaluations

Installation, Equipment, Maintenance

The InfoTrac system arrived at Brown for a 6 week trial period, beginning on April 23, 1986. It stayed for nearly 3 months. A multi-station system was sent, but only 1 station was set up due to the distance separating the 2 possible locations (Documents and the Periodicals Reading Room). Although Information Access personnel indicated that this presented no great difficulty, snaking a cable through the ceiling from the north side of the building to the south side was not something that could be easily performed by the Serials staff. Thus the location for InfoTrac became the Periodicals Reading Room.

The system arrived in approximately 15 boxes and all were well labeled as to their contents. Storing the boxes presented a problem, though, as they took up quite a bit of space. An instruction manual was included, but the graphics in it were fuzzy and during assembly it was often difficult to determine from the manual where a particular part was to be placed. The manual was written as though it were for a prototype; not much information had been added to assist librarians who might be setting the system up for the first time. Installation was sort of a seat of the pants affair, but it was managed.

In addition to the problems with the manual, installation was complicated by an equipment failure. Once the system was assembled, an error report indicated a problem with the video disk player; it failed to come up to speed. Fortunately, since the Library was sent a multi-station system, the second video player could be substituted for the first. The second player did work and the system became operational. (Once Information Access was informed of the equipment malfunction, they quickly replaced the defective player).

Two other equipment related isues needed to be solved at the outset. First, the noise of the fan in the machine cabinet (housing the video disk player and the multi-host interface) was extremely annoying to the staff in the Serials Workroom. The cabinet had to be shifted to a position as far from the Workroom door as possible, and also the doors in the hallway closed in an effort to reduce the noise. Still the noise permeated the office. The second matter concerned whether or not to leave the machine on all night or shut it down and restart it each morning. The decision was made to leave it on 24 hours a day. Since this was a trial, we felt that it would be good to test the system's reliability.

During the 3 months that InfoTrac was at Brown, the system was, for the most part, up and running. The only other major equipment problem that had to be dealt with was the failure of a video monitor. This problem was solved in the same way that the video player breakdown was - the monitor intended for the second station which we never set up was substituted for the defective monitor. The downtime for this repair was minimal. The most significant daily irritant was printer misfeeds, which when they occurred, caused the whole system to go down. This required the intervention of a staff member to correct the misfeed and restart the system (neither difficult nor time consuming, but annoying).

As far as routine maintenance went, the printer inkjets were checked every day and perhaps as many as half a dozen needed to be replaced during the trial. Also, the arrvial of new disks necessitated bringing down the system for installation. The total process took 15 minutes and presented no big problem. Outside of these routine matters, the staff in the Serials Workroom did not have to spend a lot of time working on the system.

Reaction to the System

Included with the InfoTrac package were several promotional posters and a stack of blank questionnaires. A poster was put up in the Periodicals Reading Room, and that represented the extent to which the system was advertised.

InfoTrac was put into place at an ideal time. It was close to the end of the semester and students were busy researching final course papers. The result was a continuous stream of users for the system. One concern in having the machine in the Periodicals Reading Room was the fear that teaching students to use InfoTrac would become a burden on the Serials staff and particularly so at night and on weekends when only student assistants were on duty. This fear proved to be unfounded. For the most part, people found the system so easy to use that little to no help was required to assist them. In fact, rather than a burden, the system proved to be of great interest to the staff and they gained great satisfaction from having it in the room and showing it off. The fact that even non-computer oriented staff went out of their way to demonstrate the system testifies to the ease with which the system can be learned.

As mentioned above, questionnaires were supplied with the system and these were placed next to the machine. After about the first week or so, a question concerning the worth of the system given its $13,000 price tag was added. The questions and the tallied results are given below:

1) InfoTrac can be used without formal training.

Agree Strongly	Agree Somewhat	Neither Agree Nor Disagree	Disagree Somewhat	Disagree Strongly
133 (91%)	12 (8%)	0	1(.6%)	0

2. Service worth $13,000 a year?

Agree Strongly	Agree Somewhat	Neither Agree Nor Disagree	Disagree Somewhat	Disagree Strongly
42 (59%)	12 (17%)	8 (11%)	6 (8%)	4 (5%)

3. Help screens provide valuable information for using InfoTrac.

Agree Strongly	Agree Somewhat	Neither Agree Nor Disagree	Disagree Somewhat	Disagree Strongly
86 (67%)	28 (22%)	10 (8%)	3 (2%)	1(1%)

4. I successfully answered my research question using InfoTrac.

Agree Strongly	Agree Somewhat	Neither Agree Nor Disagree	Disagree Somewhat	Disagree Strongly
94 (64%)	36 (24%)	4 (3%)	6 (4%)	7 (5%)

5. Overall, InfoTrac is easy to use.

Agree Strongly	Agree Somewhat	Neither Agree Nor Disagree	Disagree Somewhat	Disagree Strongly
107 (88%)	13 (11%)	0	0	1(1%)

6. I prefer to use InfoTrac over comparable reference tools.

Agree Strongly	Agree Somewhat	Neither Agree Nor Disagree	Disagree Somewhat	Disagree Strongly
80 (68%)	24 (20%)	6 (5%)	5 (4%)	3 (3%)

The results of the questionnaires support the contention that the system is easy to use and requires no formal training. An overwhelming majority of the users responded that they agreed strongly with the statement that using InfoTrac required no formal training and a similar number said that it was easy to use. A total of 88% of the respondents agreed, to a greater or lesser degree, that their research questions were answered using InfoTrac. Similarly, 88% preferred InfoTrac over comparable reference tools. The question about the worth of the system (given the hefty price tag) received a favorable reponse from 76% of the responding users.

A sample of some of the comments that users made are listed below:

1. This is the greatest thing since sliced bread! Easy to use, clear and comprehensive.

2. The Rock needs at least 7 of these. They are great.

3. I think that this is an amazing tool. I am very skeptical about computers and education, but this is one application I fully support.

4. InfoTrac found many sources for my topic and related topics in a quick and effective manner, more so than I could ever do. It was great!

5. Like the sign says, it is amazing. But it's also amazingly expensive. I'm not sure if the time saved warrants the costs. For $13,000 we could get 600 new books.

6. This thing is my life.

By and large, the comments on the questionaires were strongly in favor of the system.

Evaluation

The InfoTrac system was obviously easy to use and popular with people that used it, but Library staff charged with the review of the experiment (representaives from Collection Development, Reference, Documents, and Serials) had concerns. First and foremost was the cost of the system -$13,000 a year for 5 years and then $9-10,000 a year after that. The cost could be reduced if the Library purchased its own equipment, but this would be risky because of the changing technology. The advantage of purchasing the machinery and the service from Information Access was that if the technology changed, they were willing to replace the old system with the necessary new machinery. Another factor in the cost was that no one felt that any existing indexing tool could be cancelled to help pay for InfoTrac. It did not really replace anything.

A second concern was that Brown owned only about one-third of the titles in the InfoTrac database and that the database was strongly slanted toward business titles (not a strong interest at Brown since the MBA program died). It was felt that this would result in a great deal of user frustration and perhaps to an increased level of interlibrary loan requests or pressure to purchase more of the titles that were in the database. The fact that the database only extended back 3 years was also considered a drawback (this was mentioned on several of the questionnaires as well).

It was also strongly felt that the caliber of the titles indexed was not scholarly enough and, as a result, it would not really service the research needs of the campus. There were very few foreign language titles for example. It was felt by some that the presence of InfoTrac would result in patrons not utilising to the best reference sources and that they would feel like they had covered the universe when in fact they had done no such thing.

Another concern was the subject searching on InfoTrac which was felt to be cumbersome. The lack of a keyword search capability was felt to be a serious drawback and the reliance on LC subject headings made zeroing in on a topic difficult.

Some members of the Reference staff had difficulty justifying the purchase of InfoTrac for the reasons in the above paragraphs. They did not feel that they could in good conscience reduce the amount of money available for the purchase of scholarly books and periodicals in order to pay for InfoTrac.

Summary

The Library found the InfoTrac trial to be a valuable experience. Some of the results were surprising and some were not. The highly enthusiastic response of the library's users to the system took everyone by surprise as did the ease with which people were able to learn and operate the system. It was a major disappointment to see that the content of the database included so much non-scholarly and business related material, and to discover that Brown owned only a small percentage of the indexed titles. It was interesting to see the amount of maintenance required and witness the downtime firsthand (neither was appreciable).

In the end, the cost of the system was the key element in the Library's decision not to retain InfoTrac. However, there were also strong philosophical differences with the shortcut approach to library research which the staff felt InfoTrac encouraged. Additionally, the content of the database did not correspond well with the University's current curriculum.

University of Illinois

Creation and Demonstration of Optical Disk Version of the University of Illinois/River Bend Library System Online Catalog

LSCA Project: III-86-0485-30

FINAL REPORT
February 16, 1987

Submitted
by
Paula D. Watson
University of Illinois Library at Urbana-Champaign

CONTENTS

I. Program Accomplishments in Relation to Objectives 1

II. Explanation of Change in Program Direction 8

III. Project Strengths and Weaknesses 10

IV. Evaluation 12

Figure 1. Evaluation Timetable 14
Figure 2. General Evaluation Plan 15
Figure 3. List of Sources and Methods Used 16
Figure 4. Perceived Economic Value of LePac 18
Figure 5. Summary of Results of LePac Expert Mode Time Tests 23

V. Conclusions and Recommendations 31

Appendix I. March, 1986. Budget Amendment 38

Appendix II. Arrangement with University of Illinois Office of Instructional Management Services for Assistance with the Evaluation Phase of the Project 44

Appendix III. Checklists and Questions Used at Sites During Group Interviews 49

Appendix IV. Questionnaires Used with UIUC Graduate School of Library and Information Science Class 64

Appendix V. User Aids Developed for the Project 68

I. PROGRAM ACCOMPLISHMENTS IN RELATIONSHIP TO OBJECTIVES

A. General Program Objectives

1. Research and demonstrate the feasibility of creating an optical disk version of the bibliographic database of the University of Illinois Library Online Catalog now in use at UIUC and at the Riverbend Library System. The optical disk-based catalog to be created was to be searchable by means of a microcomputer using keywords in at least as many fields as are now keyword searchable in the present UIUC/Riverbend Online Catalog. The catalog was also to serve as the prototype for an optical disk version of the ILLINET Online Union Catalog now under development. Testing of the feasibility was to include an evaluation of the timeframe for initial mastering as well as updating the catalog.

2. Test the proposition that distribution of an optical disk version of the statewide ILLINET Online Union Catalog would be an effective means of enhancing resource sharing particularily for those libraries in the state which do not want or cannot afford the telecommunications cost of access to the mainframe system. Testing of this proposition was to include evaluating the utility and effectiveness of the optical disk version of the Online Catalog by observing and questioning library staff and patrons who used it.

3. Investigate whether the equipment to be purchased for use with the optical disk version of the statewide online union catalog could be used effectively for other purposes in the libraries where it might be located.

B. Program Accomplishments

1. Research and demonstrate the feasibility of creating an optical disk version of the UIUC Online Catalog database.

Due to the general advancement of the optical information industry and a continued development effort in applications for libraries, the feasibility of creating an optical disk version of a database the size of the UIUC Online Catalog is no longer at issue. Of course, when the original grant proposal was written this was not the case. There were, at that time, no packaged public access catalog products available, and our initial plan was to create a customized system to fulfill the goals of the grant. Arrangements were made with LaserData to build a system on 12-inch videodisc using the hardware employed for InfoTrac with BRS Microsearch as the search software. As has been explained in earlier quarterly reports and will be reiterated under Section III below, considerable difficulties were encountered working with a forefront entrepreneurial company whose long-term business interests did not necessarily lie in the creation of catalog products for libraries.

By the time it became clear that LaserData would not be able to make good on its original, informal agreements with us, Brodart, a well-known library products company, had entered the market with a public access catalog system. The system was still being refined and developed but it was in production. Once Brodart was identified as a vendor, the feasibility of creating an optical disk version of the UIUC Online Catalog was no longer really in doubt. What was uncertain were the ultimate retrieval capabilities of the system and the size of the database that could be created.

The first objective of the grant has been accomplished. Using Brodart's LePac system, a slightly reduced version of the UIUC Online Catalog database (700,000 records) has been put on optical disk. The database is searchable by means of a microcomputer using Brodart's proprietary search software. Searching is done in two modes: novice and expert. Had we produced a system which was searchable only through use of the novice mode we would not have fully met this first objective of the grant, since the novice mode is not geared to keyword searching. In novice mode the catalog may be searched by author, title and subject. Prompts on the left-hand side of the microcomputer screen lead the user through a quite simplistic search protocol. Alphabetical lists of either authors, titles, or subjects appear on the right side of the screen. The alphabetical lists run through the entire alphabet and prompts bring the user, through a series of passes, closer and closer to the section of the alphabet in which the item desired will be found. Through the use of "up" and "down" keys the user can isolate what he wants and then bring up a full bibliographic display for the item.

The expert mode provides the search capabilities specified in objective #1 of the grant proposal. It allows for keyword searching in all the fields where keyword searching is possible in the UIUC Online Catalog itself as well as keyword subject searching. The software also provides an "anyword" search key which permits a search for a term anywhere in the bibliographic record. Some boolean searching is possible, but truncation is not an option.

Objective #1 of the grant also specified demonstrating that the optical disk version of the database could be created within the original timeframe of the grant, i.e., in less than one year from July 1, 1985. The various factors which caused the timeframe to slip during the course of the grant period are explained in more detail under Section III below. While these factors were beyond the control of the investigator, they do indicate that during the original grant period, July 1, 1985 to June 30, 1986 it would not have been possible to accomplish the goals of the grant. Several developments occurred during the grant year which now make it quite feasible to create an optical disk-based catalog database in a time considerably shorter than a year. First, during 1985-1986 mastering facilities for CD-ROMs were introduced into the United States. Before mastering facilities were available in this country, data had to be sent abroad to Japan or the Netherlands for mastering, thus greatly increasing production and update time. Brodart began using 3M mastering facilities in early 1986. A review of the timeframe for creating the optical disk appears below.

July-August, 1985:	Continued planning discussions with LaserData which were begun before the grant was awarded.
August-October,1985:	Total absence of communication from LaserData despite repeated attempts on the part of the principal investigator to make contact.
October 30, 1985:	Complete breakdown of arrangements LaserData.
January, 1986:	Price quote and agreement with Brodart.
May-September, 1986:	Mastering of disk.

The experience with Brodart clearly indicates that it is possible to place even a large bibliographic database on CD-ROM within a reasonable period of time. For us in this instance the process took eight months, but could have been accomplished in a much shorter time had Brodart been expecting to have to fit us into their production schedule and had there been no delays at our end. (Minor delays at our end were caused by demands on programmer and machine time which prevented immediate unloading of the tapes. Also tape decks had to be identified for the purpose. To save costs, tapes which held obsolete and erasable data were used.)

2. Test the proposition that distribution of an optical disk version of the statewide ILLINET Online Union Catalog would be an effective means of enhancing resource sharing.

To positively assert that the availability of a CD-ROM union catalog would enhance resource sharing requires the examination of exactly where such a system fits into the array of resource sharing alternatives already available in a state where library cooperation has received priority attention for decades. ILLINET already provides many of the mechanisms for effective sharing of resources.

The LCS system, which links more than 25 academic and special libraries in the state and which can be accessed by the system libraries, already provides direct user access to a very large number of holdings. Some non-LCS libraries (three of the sites where the CD-ROM catalog was tested), for example, can easily gain access to LCS through dial-up and use it fairly routinely to identify material for loan.

Based on this study it would appear that a CD-ROM based catalog would have most to offer to smaller libraries in the state (public, school, college and junior college) which do not have access to OCLC, ready access to LCS, or, in some instances, for the very small libraries, access to basic identification and verification tools such as Books in Print. Obviously, such libraries, would be least able to afford the basic equipment investment or the database creation costs, depending on how these were amortized.

Paul Johnson, of the Rolling Prairies Library System, estimates that 80% of the interlibrary loan requests received by systems headquarters come from public and school libraries, many of which are without the basic bibliographic services mentioned above. Approximately 40% of loan requests come without verification. Widespread distribution of a CD-ROM catalogs would allow the smaller library to specify more accurately what it wanted to borrow. These new abilities, if properly utilized, would greatly change the character of bibliographic work done at the system libraries and, presumably, would greatly reduce the interlibrary loan verification workload.

The larger libraries in the test did not necessarily see a CD-ROM catalog based on ILLINET Online Union Catalog tapes as being very useful in the interlibrary loan process. By its nature, a CD-ROM catalog cannot have status information. A CD-ROM ILLINET online union catalog could easily have holdings information specifying system, individual library and even branch, since the information is stored in the 984 field of the record. Presumably, this information could be added to the disk and made part of the display as part of the mastering process. Due to the construction of the FBR files, the records do not have call numbers and the CD-ROM terminal did not have charging capability. A future system could in fact be configured so as to have charging capabilities, since links to the various circulation systems in use in the state are technologically feasible. Larger test site libraries did not show much interest in the addition of a circulation capability, since interlibrary loan channels already available to them (the regional system, OCLC and LCS) seem completely satisfactory.

The larger test site libraries also did not seem place a very high value on the CD-ROM catalog as an end-user tool. They tended to see it more as an aid, and a not very good one, compared to available methods, to interlibrary loan. The LePac system, which is designed as a public access catalog, is highly user friendly. There is a school of thought that public library users are mainly interested in what is available to them on the shelf in their own libraries. Thomas Ballard contends that patrons are "unwilling to accept a definition of access to books that requires them to go somewhere else or to await future delivery."[1] On the other side of this argument we have Illinoisan William Gray Potter who has pointed to the "great untapped demand for resource sharing" revealed by the LCS experience and the importance of finding ways to "make access to the combined collections of library networks as easy, direct and painless as possible".[2] At both Decatur Public and Parkland College, librarians encouraged public use by strategic and highly visible placement in the library and by frequent referrals of users to the system with an explanation of its capabilities. The effort at Decatur bore fruit in an approximately 15% increase in interlibrary loan activity for the test period. About 5% of this was accounted for by the particular enthusiasm and subject interests of two members of the staff. The night janitor, who asked to have the system turned on during her 11:30 p.m lunch hour, had a strong interest in genealogy and computers which resulted in the loan of on the order of a dozen books. A reference librarian also had a strong subject interest and he borrowed several books from UIUC which could be identified by a LePac search. The remaining increase was accounted for by members of the general public identifying books of interest on such topics as "bridges", for example, or on "cranes" and "derricks". Enthusiasm for the system was great at Parkland too, but there is no measurable increase in interlibrary borrowing to report. It is possible that Parkland users went directly to the UIUC Library to obtain materials identified through LePac. However, there is no evidence to substantiate this supposition one way or the other.

An element of the feasibility of distributing the ILLINET Online Union Catalog database on optical disk which must still be discussed is the technical feasibility of mastering and searching the full 4,000,000-record database as opposed to the trial 700,000-record database. (Only 700,000 records were put in optical format, incidentally,to avoid the necessity of employing two disks and to keep within the dollar amount budgeted for this purpose in the original grant proposal). This issue will be discussed more completely under Conclusions and Recommendations.

A third element of the accomplishment of program goal #2 is the evaluation of the utility and effectiveness of the optical disk version of the Online Catalog. This work is described in detail under Evaluation (Part IV) of this report.

One of the unexpected results of this demonstration project was that Parkland's Director, David Johnson, has become seriously interested in a CD-ROM public access catalog for his library. Staff at Parkland are very dissatisfied with the CLSI system provided to them through Lincoln Trail Library System. A CD-ROM catalog would provide them with a cost-effective alternative. There are several CD-ROM public access catalog products coming into the marketplace now. Parkland is evaluating them and obtaining cost figures.

To summarize the work accomplished in relation to program goal #2, we believe we have shown that CD-ROM based catalogs in the right locations can enhance resource sharing. Issues of location and the scope of potential databases are discussed further in the Conclusions and Recommendations section below. While the projected size of the ILLINET Online Union Catalog database is a potential barrier to distribution on CD-ROM, technological developments may occur with sufficient rapidity to remove size as a concern. Compromises on the amount of data included and/or other manipulations of the data might still make possible searching and mastering of the whole database with today's technology.

3. Investigate whether the equipment to be purchased for use with the optical disk version of the statewide online union catalog can be used effectively for other purposes.

If the CD-ROM based catalog is to be used, as the investigator believes it should, primarily as an end-user tool, it should probably be set up continuously to provide information on the holdings of other libraries and for no other purpose. More than one purpose could be served by changing software and disk and searching other optical disk-based systems, e.g., reference tools, indexes and abstracts. The user would find this confusing and it would vitiate the effective use both of the CD-ROM catalog as well as any CD-ROM reference tools which might be purchased. Optical industry experts believe it is going to be hard to develop reference tools for public libraries. As Parke Lightbowne remarked at the pre-conference preceding the ALA Conference in July, 1985, "public libraries will be hardest to supply with optical sources because concentration will not produce cost effectiveness. . . ." Few general reference sources that are at all within the budgets of public libraries are now being sold, and, in fact most development has been in more subject-specialized indexes in the sciences and the social sciences. The Grolier Academic Encyclopedia is a possibility, Bowker's Books in Print product and the brand-new InfoTrac II are also candidates for public library use. InfoTrac II, a lower-priced, CD-ROM spin-off of the 12"-videodisc-based InfoTrac I has been specifically designed for the public library market. Even were some kind of jukebox arrangement perfected for CD-ROM's, it still seems hard to imagine having the same equipment used for many purposes unless some interface software were employed to guide the user from a search from one radically difference form of information (books outside the library) to another (articles, facts).

II. EXPLANATION OF CHANGE IN PROGRAM DIRECTION (See Appendix II. Budget Amendment)

1. Original Program Plan

The original proposal named LaserData of Cambridge, Massachusetts as system integrator. We intended to use the same hardware that the company had developed for Information Access Company's InfoTrac system. The hardware was to include an IBM AT, a Pioneer LDV1000 videodisc player, a LaserData TRIO 110 controller with an interface card and a Panasonic TR 124 monochrome video display monitor. LaserData was to preprocess the data and to integrate the system. The search software to be employed was to be BRS/SEARCH. We had reservations from the start concerning the use of BRS/SEARCH which runs on the UNIX operating system, a system well known to present difficulties for users. During early negotiations with LaserData, we asked them to develop or identify a DOS-based software that could be used on the project. They seemed initially interested in doing so and claimed to have already begun work in this area.

There was essentially no contact with LaserData from September 1, 1985 to October 31, 1985. It was only through the intervention of an individual in the industry that contact was reestablished. At that point LaserData made clear that it was not willing to produce a product in any way close to the one that had originally been agreed upon. They were prepared to master only a 10 megabyte database (a very small percentage of the UIUC database) and they were not prepared to identify an MS-DOS based search software.

After further consultations with individuals in the industry, Brodart was identified as a possible supplier. The company had recently begun to market a CD-ROM public access catalog and were refining it at a rapid rate. We agreed to work with Brodart after meeting with representatives of the company at ALA Midwinter in January, 1986. The project timeframe thus slipped six months and we found ourselves working with CD-ROM rather than with 12" videodisc. Equipment requirements were entirely different. Cost changes are reflected in the budget amendment document in Appendix I.

2. Changes in the Evaluation Plan

The original evaluation plan called for direct comparison of searches on the UIUC Online Catalog with searches on the CD-ROM-based version of the Catalog. Once the LePac system was in place, it became clear that its search capabilities were so far inferior to those of the mainframe-based system as to make direct comparison of searches not very useful. Due to the slippage in the grant schedule, it was also decided to contract out for assistance with the evaluation of the system. Accordingly, an

agreement was made with the University of Illinois, Office of Instructional and Management Services, Division of Measurement and Evaluation to work with the principal investigator on the evaluation phase of the project. Dale Brandenburg, Associate Head of the Division was assigned by IMS as the project evaluation officer. See Appendix II. The evaluation plan is described in detail in Part IV of this report.

III. PROJECT STRENGTHS AND WEAKNESSES

1. Strengths

The major strong point of this project is that it introduced and demonstrated a new technology at a point of high interest among librarians in the state. The result of this timeliness was to generate or reinforce significant interest in the implementation of a CD-ROM catalog at Parkland College. Parkland has begun the study of alternate systems and sees itself on something like a one-year timeframe to install a CD-ROM-based catalog. The Director and staff see numerous benefits from this course of action. They are currently quite dissatisfied with the online circulation system in use throughout the Lincoln Trail Library System and find that it does not at all meet their needs. A product like LePac would give them cost-effective control over their own database and, they believe, good access for their users.

2. Weaknesses

The chief weakness of the project was its failure to demonstrate the technology in a realistic setting. Sites which could not be expected to have direct access to the ILLINET Online Union Catalog ought to have been chosen to evaluate the system. Best results in evaluation were obtained at Decatur which has no direct access to LCS and which is only beginning to use OCLC for cataloging. The only exposure smaller libraries received to the system was at the Reference Fair held at Lincoln Trail Library System headquarters on November 21, 1986. Ann Neely, of Parkland College, demonstrated its use there and reported an enthusiastic response from the librarians from small public libraries and high school libraries who were present. This is, of course, only anecdotal evidence of positive interest, rather than a test of use in an actual setting. Particular interest was shown by the librarian from Marshall High School, who has made it a practice whenever possible to expose the students to resources at large libraries by scheduling research trips to nearby universities. The librarians who saw the system were naturally concerned with costs and also with updating schedules.

Another weakness in the principal investigator's view was the short time available for evaluation. Because of the time limitations, only a small amount of data could be realistically collected from users and this prevented exploring various areas in more detail. For example, it would have been interesting to query the users as to the level of their interest in access to remote library holdings.

The UIUC Library database was chosen for the demonstration for obvious reasons of convenience. It might have been useful to create a different database for testing, for example, one consisting of the holdings of LCS schools in a particular, limited geographic area to be tested at small institutions in that area.

V. CONCLUSIONS AND RECOMMENDATIONS

1. Current Technical Feasibility of Placing the ILLINET Online Union Catalog Database, as Opposed to the UIUC Database, on Optical Disk.

Storage Capacity. Brodart reports that it would be just possible to fit a database of 3 million records on four CD-ROM disks. At present, most vendors claim that four drives in a daisy chain configuration are the maximum number of drives that can be strung together and searched by the same software. However, a breakthrough in the technology may be just about to occur or may indeed already have occurred. While only one month ago Brodart reported that Hitachi was working on stringing eight drives together, they now report fair confidence that they can work with Hitachi in creating a system that would run on sixteen drives. Seemingly far ahead even of these developments, Online Computer Systems announced at the American Library Association Midwinter meeting in January, 1987 that, through its Optinet network, access can be provided to 32 drives from a local area network host. It is hard to imagine such an arrangement in a small public library and the cost would, in any case, be prohibitive. Online Computer Systems quotes a price of $1800 for only two drives and the controller. At any rate, as indicated above, most vendors have reasonable confidence that they can work with a database needing four drives.

Thinking in terms of the ILLINET Online Union Catalog database, we are dealing with at least three million records, or possibly four million records, since three million was the number quoted when the system was first proposed. Mastering the initial disk would therefore be approaching what appear to be the present limits of development although not of potential. There are ways to reduce the amount of data on the disk to obtain the maximum number of records, for example, truncating the records or reducing the indexing available. On the other hand, the addition of location data to the indexing would require more space. (The project disk had neither call numbers nor locations). The union catalog has been projected to grow at the rate of 15 percent each year so that even if the initial database could be successfully mastered, updates would push the limits of development even further. Again, cost and convenience become issues. CD-ROM drives are shrinking in size and coming down in price, but current pricing would call for an expenditure of at least $2400 per site for the CD-ROM drives alone.

Furthermore, vendors have not had a great deal of experience with large bibliographic databases. At this writing, Brodart will have mastered its largest bibliographic database: 1.3 million records for Access Pennsylvania. The database will be on two disks searchable from the same software with the user unaware that both disks are being searched. General Research Corporation has a four million record database on disk for its LaserGuide acquisitions system, but access points and search capabilities are more limited than would be expected for a public access catalog.

Response Time: Experience, together with the time tests done on the LePac as part of the project, engender some concern that response time on a 4-million record database might be problematic. According to Brodart, its software version 4.9, which is to become the standard for LePac, provides five times faster response than the version we used in the demonstration. Speed of response is a well known limit of CD-ROM technology because of the high seek time required to locate data on the disks as they are presently packed. McQueen and Boss quote an average seek time of 500 milliseconds to one second with worst case access time of four seconds. This is substantially slower than magnetic media. According to McQueen and Boss, "Advocates of the technologies suggest that these speeds can be significantly improved by attention to the layout of the data on the discs and by development of specialized software. . . .[3] In fact, Brodart reports that the Hitachi drives they will use in the future contain a newly developed chip which improves seek time. The data transfer rate for CD-ROMs is also relatively slow. McQueen and Boss point out that improvements are possible here too but they require the development of faster sensors and circuits. These developments, unlike other improvements in CD-ROM technology, will not be driven by the mass consumer market for audio CD's and will therefore lead to higher equipment prices for CD-ROMs.

Newly Available Systems. As has already been pointed out, when we began this project there were no optical public access catalog products on the market. During the year since we made contact with Brodart, several have appeared and about a half dozen were demonstrated at the the Midwinter meeting in January. The principal investigator had the opportunity to examine in detail only three, LaserGuide from General Research, Spectrum from LSSI/Gaylord and Marcive's PAC. LaserGuide claimed a 3 second response time and had some key features which are missing and evidently not planned for LePac. The system appeared to deliver one match in response to a search more or less instanteously, showed a continuous number of matches in a window on the screen as they were identified, and continued to produce matches at what seemed like a quite rapid rate. The system also had a very simple approach to Boolean searching employing three windows with very general instructions concerning how to use "and", "or", and "not" operators without confusing the user by calling them "and", "or" and "not" operators. Marcive demonstrated a write-once system with some very nice features, but this product is quite limited in the size of the database it can handle. The Western Library Network has also designed a CD-ROM catalog which could be made available to licensees. This was not demonstrated at ALA and at least one critic suggests that it may be a little too sophisticated for the average public library user: "As an access tool in the average public library Lasercat still needs some work. Too many options are on the screen at once. The concepts, both computer and library, are not sufficiently transparent to make the first-time public library user comfortable. In the academic environment, where motivation to conduct exhaustive searches is greater and exposure to computer concepts more common, the current set-up should work better. A friendly if somewhat dumber novice mode would make sense, with more advanced features accessible to staff and to users who wish to get it."[4]

RECOMMENDATION #1: IN LIGHT OF THE CURRENT STATE OF THE ART, THERE MAY BE SOME SERIOUS PRACTICAL PROBLEMS INVOLVED IN PLACING A DATABASE THE SIZE OF THE ILLINET UNION CATALOG ON CD-ROM AND OBTAINING REASONABLE SEARCH AND RETRIEVAL CAPABILITIES. The developmental time frames alluded to in this report, however, indicate that the technical barriers which may currently appear to exist in relation to the creation and manipulation of large databases may be minimized within not too long a period of time. Both technological and product developments should be carefully monitored. Also, contact should be maintained with library systems currently experimenting with growing databases on CD-ROM. The Access Pennsylvania project, as mentioned above is moving to a two-disk database of 1.3 million records. Brodart reports a consortium of Long Island, New York libraries who are building a large database. They also report that the State of Wisconsin is contemplating creating a database of 4 to 6 million records within the year. Other states which are contemplating or developing statewide CD-ROM networks in place of online systems include Louisiana, Mississippi and Maine.

2. Perceived Needs / Interest/ Value

Interest. The project, unfortunately, cannot be said to have demonstrated a strong interest on the part of the site librarians in making bibliographic information on statewide holdings directly available to their users. They tended to see the optical disk catalog more as tool for librarian-mediated interlibrary loan which is far less efficient than tools currently available to them. There was a recognition that patrons could benefit from having information on volumes not available in their libraries, but there was not a sense that the need was critical. There was general agreement that smaller libraries could benefit most but could afford it least.

Costs. An informal price quote from Brodart indicates that initial mastering of the statewide database would cost $25,000 with an additional $30,000 for authority control, or $55,000 total. Additional disks would cost $30 each in quantities of 50 to 249. Updates adding all absolutely new records would cost from $27,000 to $45,000. Amortizing the cost over, for example, 300 libraries would mean each library would pay about $500 dollars during the first year for the database with two updates. Then, of course, there would be the cost of the equipment. Assuming current prices, the cost for each site would be about $4,000 for a microcomputer and four CD-ROM drives, presupposing four would be sufficient.

Value. The project queried only librarians concerning the value of a CD-ROM based statewide union catalog. Users, who have the most to gain, were not questioned and, indeed, to try to do so might be an impossible task. The experience of librarians demonstrates that users rarely understand the value of a service until they have it.

To aid in speculating concerning the potential value of largescale direct patron access in smaller libraries to statewide holdings, it might be useful to consider the experience gained in the Access Pennsylvania project. This project began as an effort to improve resource sharing among high school libraries. The first step was to convert high school library catalogs to machine-readable form. Brodart provided the most attractive bid and got the contract. At this point, no one had any idea that a CD-ROM union catalog would be created. LSCA Title III funds were obtained to support the project which required multitype library networking. Eighteen or nineteen consortia were formed around the state which consisted of high school libraries, public libraries of various sizes (Carnegie Library of Pittsburgh with 300,000 volumes is the largest) and smaller university libraries. Most libraries in the consortia are high school libraries, but some of the participants do have access to OCLC.

While project developers have not done detailed analysis of the flow of materials, the experiment has had some gratifying, documented results. For example, an isolated high school library, fifty-two miles from the nearest public library has quadrupled its own circulation because students have automated access to the 8,000 volume collection. Students also borrow with some regularity from a small university library in the consortium.

High school circulation has greatly increased. The general flow of materials within the consortia has also increased and, project developers believe, though they have not documented it, more materials are also circulating from one end of the state to the other. These increases in circulation may certainly be seen as evidence of the value of the project.

Potential Illinois Applications of CD-ROM with Smaller, More Targeted Databases.

Phasing in the ILLINET Online Union Catalog. Since all potential participants in the online union catalog will not be able to join immediately, a CD-ROM based system might serve as an interim solution. The database could consist of either the complete holdings of the, as yet, non-participant libraries, or regional subgroupings of holdings or of LCS/FBR holdings or subgroups of LCS/FBR holdings. Location information can be indexed on the CD-ROM and, for non-LCS libraries, call number information could be provided. Alternatively, the CD-ROM catalog could be interfaced with existing circulation systems to make a full-featured online catalog.

Individual Libraries on CD-ROM. Like Parkland, any library in the state which is not linked to an online system or is not satisfied with the online system it is linked to might be encouraged to look at a CD-ROM system as an alternative to an online catalog. This idea has obvious appeal for Decatur Public Library patrons. The systems now available are ideally suited for individual libraries with collections of about 100,000 volumes or small groupings of such libraries. Total local control is possible, the systems are cost effective, and quite sophisticated search capabilities with reasonable response times can be provided in a highly user-friendly presentation.

Extend Automated Access Beyond Present Boundaries. Following the Pennsylvania model, consideration might be given to extending automated public catalog access on CD-ROM to libraries, for example, high school libraries, which do not presently have it. A program of this kind is also underway in Minnesota.

RECOMMENDATION # 2. FURTHER EVALUATE THE LEVEL OF SMALL LIBRARY INTEREST AND OF PATRON INTEREST IN LARGER LIBRARY COMMUNITIES IN DIRECT ACCESS TO THE HOLDINGS OF OTHER LIBRARIES. CONSIDER A CD-ROM CATALOG WITH A CIRCULATION MODULE AS AN INTERIM SOLUTION FOR LIBRARIES AWAITING INCLUSION IN THE STATEWIDE CATALOG. CONSIDER THE POSSIBILITIES FOR SMALLER LIBRARIES WANTING AUTOMATED ACCESS TO THEIR OWN HOLDINGS BUT WHICH ARE UNABLE TO AFFORD A FULLY ONLINE SYSTEM OR WHICH ARE NOT LIKELY TO BE SATISFIED WITH WHAT CAN BE MADE AVAILABLE TO THEM.

FOOTNOTES

1. Thomas H. Ballard, "Dogma Clouds the Facts," American Libraries, 16 (April 1985): pp. 258-259.

2. William Gray Potter, "Creative Automation Boost ILL. Rates," American Libraries, 17 (April 1986), pp. 244-246.

3. Judy McQueen and Richard W. Boss. Videodisc and Optical Disk Technologies and Their Applications in Libraries. 1986 Update. (Chicago: American Library Association, 1986), p. 64.

4. Karl Beiser, "Microcomputing," Wilson Library Bulletin (November, 1986), pp. 41-42.

Vanderbilt University Library

ENHANCED INFORMATION ACCESS PROJECT

Progress Report #2
6 February 1987

Caravan Network Extension

Cabling has been completed for the anticipated network connections to the Medical, Central, and Science libraries. The broadband is in place and design specifications for in-building wiring are being prepared. IBM and DEC components necessary for allowing dial-up and network access to ACORN and other services supported on the Library's IBM have been ordered. Delivery and installation of these components, which will support 32 ports, are expected by March.

Article Level Files - CD-ROM Experiments

In accordance with the agreement signed with OCLC, the Education Library has served as a test site for the CD-ROM product of CIJE. Two workstations were installed in mid-August and have been used continuously since then. Surveys collected from library users and suggestions from the librarians have resulted in several modifications in OCLC's product. OCLC has publicized Vanderbilt's role in this developmental project and an article by a member of the Heard Library staff describing the Education Library's experience with this experimental product is currently under review for publication.

The Science Library housed OCLC's Graph=Text system for about sixty days. This test product consisted of a CD-ROM containing the full text of chemical journals, produced in cooperation with the American Chemical Society. OCLC representatives have indicated that they are still hoping to provide portions of MEDLINE and bibliographic records for government documents in CD-ROM formats and are anxious to have us continue as a test site, but no firm dates have been set for these projects. Their representatives are making contacts with various government agencies to enable development of these products.

Other experiments with compact optical discs are underway and planned. In addition to the OCLC version of CIJE, the Education Library has obtained SilverPlatter's CD-ROM version of the ERIC database and staff are monitoring user response to and comparisons of the two implementations. Staff members have experimented with Dissertation Abstracts Ondisc, but because of a temporary lack of availability of equipment, this disc has not been made widely available to the public. Negotiations with Information Access Company have led to a test of InfoTrac in the Central Library. This system will be installed in the Reference room the week of February 9 and will be available throughout March and April. Le Pac: Government Documents Option from Brodart has also been ordered and will be installed in the Documents unit at Central this month. The Management Library conducted a 30-day test of Datext's CD/Corporate compact disc product and, due to an overwhelming vote of approval from its users, elected to continue the subscription for at least six months. Our desire to implement network access to CD-ROM products has been discussed with several vendors doing developmental work in this new field and appears to be of interest.

Vanderbilt University Library

Article Level Files - Cooperation with NOTIS

Our objective of loading databases on magnetic tape on our local computers and providing for searching through the network was the focus of a meeting on January 9 with representatives of NOTIS and key Vanderbilt personnel. A preliminary requirements document prepared by NOTIS is attached. Contract negotiations with NOTIS and specifications for the NOTIS linkage product are currently under review and should be finalized by the end of February. We anticipate that this cooperative effort should provide a smooth transition for the library users to searching article-level databases with techniques similar to those used in the online library catalog and will eventually be of benefit to other NOTIS users as well as to Vanderbilt. Our plan is to begin this project with the loading of a MEDLINE subset (probably the last two years). Discussions are in progress with vendors of other online databases, but no other commitments have been made at this time.

Selected Reading List

SELECTED READING LIST

- Bain, Alan L. (Associate Archivist, Smithsonian Institution Archives), "Conversion of Archival Materials to a Binary Format," 35p., draft article.

- Lambert, Steve and Suzanne Ropiequet, eds. The New Papyrus. Redmond, Washington: Microsoft Press, 1986.

- McQueen, Judy and Richard W. Boss. Videodisc and Optical Digital Disk Technologies and Their Applications in Libraries, 1986 Update. Chicago: American Library Association, 1986.

- Miller, David C. Four Trend Reports for the Fred Meyer Charitable Trust, Library and Information Resources for the Northwest:
    - Laser Disks Knock on the Library Door (3/86)
    - The New Optical Media Mid-1986: A Status Report (8/86)
    - The New Optical Media in the Library and the Academy Tomorrow (8/86)
    - Moving Information: Graphic Images on CD ROM (11/86)

    Available from David C. Miller, Managing Partner, DCM Associates, Post Drawer 605, Benicia, CA 94510 (707/746-6728)

- Ropiequet, Suzanne, ed. CD ROM 2: Optical Publishing. Redmond, Washington: Microsoft Press, 1987.

- Roth, Judith Paris. Essential Guide to CD-ROM. Westport, Connecticut: Meckler, 1986.

- Saffady, William. Optical Disks for Data and Document Storage. Westport, Connecticut: Meckler, 1986.